Summer 6-22-2017

N400, P600, and Late Sustained Frontal Positivity Event-Related Brain Potentials Reflect Distinct Psycholinguistic Processes During the Comprehension of Contextually Ambiguous Narrative Discourses: Implications for the Neural Organization of the Language Processing System

Patrick S. Ledwidge
University of Nebraska-Lincoln, patrick.ledwidge15@gmail.com

N400, P600, AND LATE SUSTAINED FRONTAL POSITIVITY EVENT-RELATED BRAIN POTENTIALS REFLECT DISTINCT PSYCHOLINGUISTIC PROCESSES DURING THE COMPREHENSION OF CONTEXTUALLY AMBIGUOUS NARRATIVE DISCOURSES: IMPLICATIONS FOR THE NEURAL ORGANIZATION OF THE LANGUAGE PROCESSING SYSTEM

by

Patrick S. Ledwidge

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychology

Under the Supervision of Professors Dennis L. Molfese and Art C. Maerlender

Lincoln, Nebraska

June, 2017

N400, P600, AND LATE SUSTAINED FRONTAL POSITIVITY EVENT-RELATED BRAIN POTENTIALS REFLECT DISTINCT PSYCHOLINGUISTIC PROCESSES DURING THE COMPREHENSION OF CONTEXTUALLY AMBIGUOUS NARRATIVE DISCOURSES: IMPLICATIONS FOR THE NEURAL ORGANIZATION OF THE LANGUAGE PROCESSING SYSTEM

Patrick S. Ledwidge, Ph.D.

University of Nebraska, 2017

Advisors: Dennis L. Molfese and Art C. Maerlender

Two studies were performed to investigate the temporal structure and organization of the language processing system during the comprehension of coherent contextually ambiguous narrative discourses. In Study 1, participants read short discourses that were contextually ambiguous if read without a descriptive title (Untitled group) or unambiguous with the title (Titled group). Participants identified the title of the discourses after reading 1-3 sentences. Given the unfinished next sentence, they performed a cloze procedure on the sentence-final word. For the Titled Group, cloze probability was greater to the last word of sentence 3 (Critical Word 3) than sentences 2 (Critical Word 2) and 1 (Critical Word 1). For the Untitled group, title identification accuracy was greater after reading the first two sentences than the first sentence and even more so after reading the full three-sentence discourses than the first two sentences alone. This study established 25 discourses in which the contexts were initially ambiguous but became increasingly clearer after reading Critical Word 2 and Critical Word 3.

In Study 2, a different sample of participants read the discourses with (Titled Discourse group) or without (Untitled Discourse group) a descriptive title while undergoing high-density event-related potential (ERP) recording. For the Untitled Discourse group, N400 amplitudes became less negative from Critical Word 1 to Critical Word 2 and again to Critical Word 3, suggesting greater ease of lexical-semantic retrieval as the discourses unfolded. For this group, P600 amplitudes were larger to Critical Word 3 than Critical Word 2. Unexpectedly, a Late Sustained Frontal Positivity (SFP) ERP component occurred to both Critical Word 2 and Critical Word 3 for the Untitled Discourse group only. SFP results corresponded to the title identification findings from Study 1. Thus, it is proposed that the SFP reflects the resolution or revision of contextual ambiguity during discourse comprehension. Alternatively, the P600 is proposed to reflect the updating of the existing context of a discourse when an context is available or after the resolution of contextual ambiguity.

Dedicated to my parents.

For your love, patience, and support.
For teaching me the value of hard work, perseverance, and integrity.

ACKNOWLEDGEMENTS

First, to Dennis Molfese, my mentor and advisor, for encouraging me to question and create— Thank you for helping me develop the skills and knowledge that are sure to guide my work for many years to come. Thank you for four years of challenges, support, and congeniality.

Second, to the rest of my committee—Art Maerlender, Karen Hux, Anne Schutte, and Ming Li—Your feedback brought innovation and focus to this project. Thank you for your time and efforts in helping me see this work through to completion.

Third, to the student-researchers who assisted me on this project—Rianne Gibson, Andrea Meinders, Sarah Bublitz, Jordan Knapp, Carly Dinnes, Jenna Varilek, Kirsten Titze, Hannah Abbott, and Jessica Salley—Thank you for all of your hard work and dedication. It was a pleasure to work with such an enthusiastic group of researchers.

Finally, to Alaina—Thank you for your unwavering support and understanding. Thank you for humoring me day after day as I rambled about "the dissertation," "the paper," "the grant," "the job, " and "the paragraphs." You gave me my most valued gifts—the ability to find balance in life and happiness in even the briefest of moments.

TABLE OF CONTENTS

TABLES AND FIGURES

CHAPTER 1

THEORIES OF LANGUAGE COMPREHENSION AND THE FUNCTIONAL SIGNIFICANCES OF THE N400 AND P600 EVENT-RELATED BRAIN POTENTIALS

1.1. General introduction to dissertation

Language comprehension is a highly interactive process. Interlocutors actively integrate new information (e.g., incoming words) with information previously presented (e.g., earlier sentences), world knowledge, and social cues to understand and convey meaning when communicating. As a result, communication is hardly just a linguistic process. One must not only attend to the ongoing presentation of phonological, lexical, and syntactical information but also integrate this information with the unraveling discourse. In essence, effective language comprehension involves syntax, sentence structure, and lexical semantics (the meanings behind words and utterances) but also the cognitive mechanisms to support the attention, integration and maintenance of this information. The functional significance and temporal organization of the psycholinguistic processes that that support narrative discourse comprehension is the focus of this dissertation.

Discourse comprehension relies on a process in which memory facilitates the match/integration of new, incoming information to the previous information presented. Theories differ in how this "matching" process occurs, but all suggest that memory plays a key role (Clark & Haviland, 1974; Kintsch, 1988; Kintsch, Patel, & Ericsson, 1999;

McKoon, Gerrig, & Greene, 1996; Myers & O'Brien, 1998; Ratcliff & McKoon, 1988). For example, according to Clark and Haviland's (1974) Given-New Strategy, this integration process fails if new incoming information (i.e., conceptual representations of words) does not match existing information in memory. In these instances, the new information will be organized in memory as a separate component that is not related to the previous knowledge. Because of this, the authors concluded that sentences containing information matching an existing mental representation are easier to understand than sentences that contradict this representation. On the other hand, sentences that contain non-matching information to the existing contextual information are difficult to comprehend.

We can apply the Given-New Strategy to situations in which the availability of contextual information influences participants' ability to understand a discourse. In their seminal study, Dooling and Lachman (1971) presented metaphorical paragraphs to participants. Some participants read a descriptive title before reading the passage, thus providing a context for the paragraph. Other participants read the paragraphs without a title. Importantly, without the title, the contexts of the paragraphs were ambiguous. Results showed that participants who read the title prior to the paragraph recalled significantly more contextually relevant words than those who did not read the title. The authors concluded that the Titled group was better able to match each individual word to the mental representation of the paragraph, as facilitated by the context, than the Untitled group. Said otherwise, the easier the integration process, the greater the likelihood of recalling words from the passages. These findings support the Given-New Strategy view: The title provided an antecedent (i.e., "context") that facilitated the matching process between each word and the mental representation of the passage (Clark & Haviland, 1974).

Bransford and Johnson (1972) also examined participants' comprehension of ambiguous paragraphs. Similar to the procedures used by Dooling and Lachman (1971), participants listened to a set of metaphorical paragraphs with or without a descriptive title. However, those who did not hear the title prior to the paragraph were given the title after the paragraph. The results showed that participants' comprehension ratings of the paragraph and the number of ideas they recalled about the paragraph were greater if they were provided with the title prior to the paragraph compared to after. In both the Dooling and Lachman (1971) and Bransford and Johnson (1972) studies, the title provided the necessary context (i.e., antecedent; Clark & Haviland, 1974) by which words could be matched to existing top-down knowledge of the discourse.

These two seminal studies indicate that the availability of a context is extremely powerful in influencing how we derive meaning from discourse. The effect of a context on discourse comprehension can be compared to a filtering mechanism; it provides the initial top-down knowledge that guides language comprehension (Kintsch, 1988). According to the Construction-Integration model, discourse processing first involves the construction of the textual information of a discourse, then the integration into the mental representation of a discourse (Kintsch, 1988).

To construct and revise a mental representation of a discourse, it is necessary for one to remember the information already presented and used to construct the existing discourse. This undoubtedly relies on working memory. Researchers suggest that working memory underlies sentence processing in which language comprehension is a cycle of encoding and retrieval processes (e.g., Lewis, Vasishth, & Van Dyke, 2006). Others also purport that working memory plays a large role in language processing (Daneman & Carpenter, 1980; Kintsch, 1988, 2005; Kintsch & Van Dijk, 1978). Although there is no denying that working memory affects language processing, the neural mechanisms involved in this process remain unknown. In fact, a pervasive debate in the field of psycholinguistics concerns the underlying neural architecture of the language comprehension system. Particularly, during language comprehension, do the two processes of (a) lexical-semantic retrieval of words and (b) the integration of these words into a contextual representation occur sequentially or in parallel? Moreover, is the integration of words into a mental model an automatic or controlled process? According to Posner and Synder's (1975) dual process model, automatic and controlled processing are functionally distinct processes. Specifically, automatic processing is uncontrolled and fast acting, whereas conscious processing is controlled, slower and of limited capacity. The purpose of this study is to further explore the organization of the underlying neural mechanisms that reflect these automatic and controlled processes during discourse comprehension.

The Resonance Model of text comprehension (Myers and O'Brien, 1998) proposes a two-step process during discourse comprehension. In brief, during reading, words activate semantic concepts in long-term memory. These concepts intersect with related concepts in working memory. Both the initial and related concepts are retrieved from long-term memory and are brought into working memory. A second process occurs when the contents in working memory are monitored and checked against new semantic concepts. This may interrupt language comprehension when a semantic concept does not match other contents in working memory. As Myers and O'Brien (1998) put forth, "Readers may refocus on the current contents of working memory, sending new signals to memory; they may engage in some form of problem solving; or they may read the next clause or sentence." The high temporal resolution of event-related potentials (ERPs) provides an optimal method for examining two-step models of discourse processing.

Specifically, ERPs are sensitive to different aspects of language processing. As reviewed in the next section, it is hypothesized that the "N400" ERP component reflects the automatic lexical process of retrieving semantic conceptual information of words from long-term memory. On the other hand, the "P600" ERP represents the post-lexical "matching" process in which the semantic elements of a word are checked/integrated against the discourse model that is currently held in working memory. A brief introduction to ERPs is provided before reviewing relevant literature regarding the N400 and P600 ERP components.

1.2. Introduction to ERPs

The present study employed high-density ERPs to examine the relative time course of automatic lexical processing and controlled post-lexical integration during discourse comprehension. ERPs are portions of the ongoing electroencephalogram (EEG) that are time locked to the onset of a sensory stimulus, motor activity, or cognitive process. It is believed that the ongoing EEG signal reflects the activity of postsynaptic local potentials from cortical pyramidal cells. ERPs reflect the aggregation of activity from these neurons. These neurons are all perpendicularly oriented to the outer layer of the cortex and "fire" synchronously to the onset of the presentation of a stimulus. Each postsynaptic potential creates a net negative charge at the location of the pyramidal neuron's dendrite and a net positive charge at the pyramidal neuron's cell body. This results in a small dipole at the location of the pyramidal neuron, in which the negative and positive electrical charges flow in opposite directions (cf. Hagoort, 2008). The ERP recorded at the scalp reflects the summation of dipole activations from pyramidal neurons that are all oriented in the same direction and fire synchronously. ERP "components," deflections within an ERP waveform, are characterized according to their polarity (positive, negative) and timing (in ms) relevant to their temporal occurrence following stimulus onset. For example, the "P100" ERP component is a positive deflection relative to a reference electrode that peaks approximately 100 ms following stimulus onset. Alternatively, the "N200" ERP is a negative deflection that peaks approximately 200 ms following stimulus presentation.

ERPs provide an optimal method for examining the time course of language processing because of their high level of temporal sensitivity. For example, EEG recorded at 1000 Hz will generate a data point for every 1 millisecond. This level of temporal resolution is an advantage of ERPs compared to other brain imaging systems, such as functional magnetic resonance imaging (fMRI) that measures blood flow throughout the brain. Blood flow is an indirect measure of neuronal activity and, as a result, maintains a much degraded temporal resolution of approximately 0.5 Hz (i.e., 2 seconds). This resolution is at least two times slower than the behavioral reaction time needed to distinguish words from nonwords (e.g., Meyer & Schvaneveldt, 1971).

However, fMRI's high degree of spatial resolution is substantially more precise than ERP; determining the neural generators of ERPs is challenging. One cannot assume that the neural origin of an ERP lies directly below the region of the scalp where the ERP is maximally recorded. This is because of the skin, skull, meninges, cerebrospinal fluid, and other structures that lie between the neuronal source(s) of the aggregated dipole and the electrode channels that sit on top of the scalp. Determining the spatial source of an ERP from the electrical potentials recorded on the scalp is referred to as the “inverse problem.” Source localization techniques estimate the source of the ERP using mathematical modeling. High-density systems for adults with spatial arrays of 128-256 electrode channels boast a smaller error in estimating neural sources compared to low-density systems with fewer electrodes (Junghöfer, Elbert, Leiderer, Berg, & Rockstroh, 1997; Song, Davey, Poulsen, Turovets, Luu, & Tucker, 2014), but this remains an imperfect solution. For example, the 256-electrode channel HydroCel Geodesic Sensor Net (ElectroGeodesics, Inc., Eugene, OR, USA) maintains a localization error distance of as low as 7mm (Song et al., 2014). Thus, fMRI remains the gold standard for spatial resolution. For the purposes of this study, ERPs were used because of their high temporal resolution.

Researchers cannot use ERPs to deterministically identify the different neural mechanisms that support automatic and controlled lexical processing. However, the relative timing, morphology, and/or scalp distribution of ERPs, coupled with controlled experimental designs, enable us to make conclusions about the relative organization and time-course of these processes in the brain (cf. Van Berkum, 2004). The goal of the present dissertation was to record ERPs elicited during a discourse comprehension task that isolated the post-lexical process by which a mental model of a discourse is updated/revised. If the ERP that reflects this process is different in timing, morphology, and/or scalp distribution from the ERP that is recorded to every single word (e.g., N400 effect; Kutas & Hillyard, 1980), then we can conclude that the neural mechanisms that underlie these different processes are, at least in part, distinct (Van Berkum, 2004).

ERPs were recorded to explore the neural mechanisms that underlie automatic and controlled language comprehension processes. This will inform our understanding of the neural architecture of the language processor. In doing so, the functional significances of two specific ERP components, the N400 and P600, were investigated. Distinguishing the cognitive processes associated with the N400 and P600 is an ongoing debate among psycholinguists. Whereas some researchers theorize that the N400 represents a post-lexical process in which a word is integrated into the mental structure of a discourse (e.g., Brown & Hagoort, 1993), others suggest that it represents an automatic process of lexical-semantic retrieval from memory (e.g., Kutas & Federmeier, 2000). The P600 is most commonly discussed as an index of syntactic revision/repair (cf. Hagoort, 2003, 2008). However, some contend that it may also represent the integrative process in which the mental model of a discourse is updated (Brouwer, Fitz, and Hoeks, 2012; Hoeks et al., 2004). To address this debate, this dissertation examined whether the lexical process that serves to update our mental model of a discourse is reflected in the N400 and/or P600 components of the ERP.

The current study is grounded in the previously reviewed hierarchical models of discourse comprehension. These include the notion that (a) lexical access/retrieval of words from long-term memory is brought into working memory and is context-independent and (b) a matching/integration process occurs in which the semantic elements of the word are checked against the current message-level representation of a discourse (Kintsch, 1988, 2005, Myers & O'Brien, 1998). To examine whether the N400 and/or P600 ERP components represents one or both of these two processes, it is necessary to first establish the moment at which individuals undoubtedly update their mental representation of a discourse, such as when they identify the context. This dissertation reports two studies that address these questions. In the first study, a novel stimulus set of paragraphs that isolated the process in which a mental representation of a discourse is updated ("discourse context updating") was developed. In the second study, ERPs were recorded while participants read these discourses to examine the extent to which the N400 and/or P600 ERP components reflect discourse context updating.

In the following sections of this chapter, relevant research is reviewed regarding the different theories of the functional significances of ERP components that support different aspects of the language processor, particularly the N400 and P600 ERPs. Chapter 2 describes Study 1; specifically, the stimulus development and validation procedures to construct the stimuli used for Study 2. Chapter 3 describes the Methods, Procedures, Data Analysis approach and Hypotheses for Study 2. Results of Study 2 are reported in Chapter 4, and a General Discussion is provided in Chapter 5.

1.3. Literature Review

1.3.1. The N400 ERP

Kutas and Hillyard (1980) first discovered the N400 ERP as a negative deflection time-locked to the onset of words that were semantically incongruous within a sentence. In their seminal study, participants read seven-word sentences one word at a time. The final word in the sentence was congruent, strongly incongruent, moderately incongruent, or visually deviant (presented in a larger letter size) within the context of the sentence. They found that a negative deflection (referenced to linked mastoids) in the 300-600 ms range was larger for the incongruent words at the end of a sentence compared to the semantically congruent words. For example, in the sentence, “He spread his bread with warm---” the amplitude of the N400 was more negative to “socks” (incongruent word) compared to “butter” (congruent word). This negative deflection began around 250 ms, peaked at approximately 400 ms, and was maximal at centroparietal electrode channels, with a slight right scalp hemisphere lateralization (Kutas & Hillyard, 1980; see Federmeier & Laszlo, 2009 for review; but see Bentin, McCarthy, & Wood, 1985 for no laterality effects). However, the visually deviant stimuli produced a P300-like effect (Sutton, Braren, Zubin, & John, 1965). The authors concluded that the N400 represented some aspect of semantic processing. Based on these and other findings, researchers theorized that the N400 represents a post-lexical process in which a word is integrated into its semantic context (e.g., Bentin, Kutas, & Hillyard, 1995; Brown & Hagoort, 1993; Chwilla, Brown, & Hagoort, 1995; Hagoort, 2008; Kutas & Hillyard, 1984; St. George, Mannes, & Hoffman, 1994).

Since Kutas and Hillyard’s (1980) identification of the N400 component, researchers conceptualized this component as reflecting a language-specific process (e.g., Holcomb & Neville, 1990). However, more recently, findings indicate that the N400 is not specific to language contexts but, rather, represents an amodal process that is sensitive to any stimulus from which “meaning” can be derived (see Federmeier & Laszlo, 2009 for review). The “typical” N400 effect, more negative peak amplitude to semantic anomalies compared to semantically coherent stimuli, is demonstrated across sensory domains.

For example, more negative N400s occur to semantically anomalous words than compared to coherent words within a sentence context that is presented both visually (e.g., Holcomb & Neville, 1990; Kutas & Hillyard, 1980) and auditorily (Holcomb & Neville, 1990; McCallum, Farmer, & Pocock, 1984). As reviewed in the following sections, the N400 effect occurs when the context prior to the anomalous word is a single word (i.e., "semantic priming task," Bentin et al., 1985; Brown & Hagoort, 1993; Chwilla et al., 1995; Deacon, Hewitt, Yang, & Nagata, 2000; Grossi, 2006; Kiefer, 2002; Rolke, Heil, Streb, & Henninghausen, 2001), a sentence (Kutas & Hillyard, 1980, 1983, 1984), and a wider multi-sentence discourse/story (Federmeier & Kutas, 1999; Fischler, Bloom, Childers, Roucos, & Perry, 1993; Nieuwland & Van Berkum, 2006; St. George et al., 2004; Van Berkum, Hagoort, & Brown, 1999). Compared to the visual N400, the auditory N400 has an earlier onset, slightly larger and wider deflection, and is more centrally distributed over the scalp (visual N400 tends to be right-lateralized, suggesting a larger involvement of the left hemisphere compared to the right; Holcomb & Neville, 1990; see Federmeier & Laszlo, 2009 for review).

In addition to the language modality, N400 effects are also recorded semantically anomalous non-linguistic stimuli such as pictures (McPherson & Holcomb, 1999), complex visual scenes (West & Holcomb, 2002), line drawings (Barrett & Rugg, 1990), faces (Barrett & Rugg, 1989; Barrett, Rugg, & Perrett, 1988; Willems, Özyürek, & Hagoort, 2008), gestures (Wu & Coulson, 2005), and environmental sounds (Orgs, Lange, Dombrowski, & Heil, 2006; Van Petten & Rheinfelder, 1995). Although it is most known for its history within the language modality, the N400 clearly reflects an amodal conceptual process because it is evident in tasks that cross sensory domains.

For example, Nigam, Hoffman, and Simons (1992) replicated Kutas and Hillyard's (1980) study with an added condition in which sentences concluded with a line drawing of the semantically anomalous words. The authors found that the morphology and timing of the N400 effect in this cross-sensory modality condition was the same as that to the typical semantically anomalous condition. Other researchers reported semantic N400 effects to anomalous picture-word combinations (Willems et al, 2008). Although timing, size, and scalp distributions of the N400 vary slightly between these modalities, this collective body of research suggests that the respective N400s may be, at least in part, supported by the same underlying neural mechanisms. This conclusion, however, must await more systematic studies using high-density electrode arrays of sufficient size to establish the veracity of such conclusions.

As previously reviewed, larger N400 amplitudes are generated to stimuli that are incongruent with modality-independent contexts compared to stimuli that are congruent with the context. Of particular relevance to the present study is the N400 recorded to words within a wider discourse. Diverging findings exist as to the extent to which discourse anomalous words generate larger N400s compared to discourse coherent words. Some researchers reported more negative N400 amplitudes when a word was incongruent within a discourse (e.g., Nieuwland & Van Berkum, 2006), whereas other researchers suggest that discourse anomalies do not produce N400 effects if the anomaly is semantically related to the theme of a discourse (i.e., the "semantic illusion phenomenon"; e.g., Hoeks et al., 2004; Nieuwland & Van Berkum, 2005). Based on these types of findings, researchers questioned the long-held belief that the N400 represents a post-lexical process in which a word is integrated into the semantic context (e.g., Bentin et al., 1995; Brown & Hagoort, 1993; Chwilla et al., 1995; Hagoort, 2008; Kutas & Hillyard, 1984; St. George et al., 1994). Rather, some researchers suggest that the N400 represents an automatic process (Brouwer et al., 2012; Brouwer & Hoeks, 2013; Burkhardt, 2006, 2007; Deacon et al., 2000; Federmeier & Kutas, 1999; Fischler et al., 1983; Grossi, 2006; Kiefer, 2002; Kiefer & Spitzer, 2000; Kutas & Federmeier, 2000; Lau, Almeida, Hines, & Poeppel, 2009; Lau, Phillips, & Poeppel, 2008; Luck, Vogel, & Shapiro, 1996; Rolke et al., 2001). In the next section, research findings on the functional significance of the N400 are reviewed.

1.3.1.1. The functional significance of the N400: A lexical or post-lexical process?

Brouwer and colleagues (2012) questioned the functional significance of the N400 as an index of semantic integration. In short, these authors provided an explanation for researchers' findings of null N400 effects to words that were semantically anomalous with a prior discourse. For example, the authors reviewed Nieuwland and Van Berkum's (2005) study which examined the N400 effect to semantically anomalous words within seven-sentence mini stories. These stories described three entities: a man and woman engaging in conversation and an inanimate object. In the carrier sentence of interest, the woman either continued to talk to the man or began talking to the inanimate object. Importantly, the inanimate object was semantically anomalous with the local carrier sentence but coherent with the context of the discourse. The authors found that anomalous words did not elicit an N400 effect. For example, N400 amplitudes were equal to the last word in each of the following two sentences: “Next, the woman told the suitcase,” (anomalous) and “Next, the woman told the tourist” (coherent). Brouwer and colleagues (2012) suggested that participants did not generate an N400 effect because the prior context of the discourse (about a tourist getting on a plane) primed the semantic activation of both the words “suitcase” and “tourist.” In fact, Nieuwland and Van Berkum (2005) found a P600 effect to the anomalous word (“suitcase”) compared to the coherent word (“tourist”). Based on these findings and others reporting P600 effects instead of N400 effects to semantically anomalous words that were lexically or contextually primed (e.g., Hoeks et al., 2004), Brouwer and colleagues (2012) proposed a novel theory to account for this and similar effects.

In brief, Brouwer and colleagues (2012) suggest that the N400 ERP represents an automatic process: the retrieval of semantic information of words from long-term memory (“lexical-semantic retrieval”; Kutas & Federmeier, 2000). This conclusion is in stark contrast to other prior work suggesting that the N400 represents a top-down process of integrating the semantics of a word with a prior context (henceforth referred to as the “semantic integration” account of the N400; e.g., Hagoort, 2003, 2008). On the other hand, Brouwer and colleagues (2012) theorized that the P600 ERP represents this semantic integration process. In their Mental Representation of what is being Communicated (MRC) hypothesis, Brouwer and colleagues (2012) suggested that the P600 represents the updating, revising, or constructing of the mental model of the “meaning” of a discourse.

According to their theory, every word within a context should elicit a biphasic N400-P600 complex in which the N400 is positively related to the ease in which lexical information of a word can be retrieved from semantic memory and the P600 is negatively related to the ease in which a word can be integrated into the mental model of a discourse. This theory coincides with Posner and Snyder's (1975) dual-process model of cognition. Specifically, the N400 and P600 ERP components represent automatic and controlled processing respectively. Importantly, the authors postulate that context effects add to the activation of semantic features of words but do not constrain the activation (Brouwer et al., 2012). In the next section, evidence for the automaticity of the N400 effect is reviewed.

1.3.1.1.1. Evidence from semantic priming, attentional blink, and dichotic listening tasks

Following its discovery, researchers conceptualized the N400 as a reflection of the integration by which a word is tied to its preceding semantic context. For example, Kutas and Hillyard (1983) modified their original design (Kutas & Hillyard, 1980) to investigate differences in N400 amplitudes to semantic anomalies and grammatical anomalies. They also extended their recording epoch compared to Kutas and Hillyard (1980) to study if a late-positive component (i.e., P300-type) followed the N400. Replicating their original results, they found that semantic anomalies elicited more negative N400 amplitudes compared to semantically congruent words both in sentence-final and sentence-intermediate positions (Kutas & Hillyard, 1983). They found no such evidence of a P300 following the N400, which led them to conclude that the N400 is not a delayed N200 as the latter is frequently followed by a P300. The authors concluded that these findings converged with their prior work (Kutas & Hillyard, 1980) and suggested that the N400 represents both automatic lexical-semantic retrieval and the post-lexical controlled integration of a word within its existing context. The grounds for the authors' conclusions are strictly conceptual—participants' access to controlled mechanisms was not manipulated. In contrast, the most convincing evidence of the functional significance of the N400 as reflecting either an automatic or controlled process comes from studies in which researchers eliminated participants' ability to exert controlled language processing.

Findings from semantic priming studies provide the majority of such evidence (e.g., Brown & Hagoort, 1993; Deacon et al., 2000; Grossi, 2006; Rolke et al., 2001). In typical semantic priming paradigms, pairs of words are individually presented. The second word in the pair ("target") is either semantically related or unrelated to the immediately preceding word ("prime"). Intermixed with these words are letter strings that are nonwords. Participants perform a "lexical-decision task" and press a button to indicate whether the string of letters is a word or nonword. A robust effect in the literature is that words preceded by a semantically related prime are identified more accurately and quickly than words unrelated to the prime (e.g., Meyer & Schvaneveldt, 1971). The prime is thought to facilitate language processing of the target if the two are semantically related. Priming is thought to result from automatic processes, controlled processes, or both (Posner & Snyder, 1975). Collins and Loftus (1975) proposed that automatic semantic priming effects occur through a process called spreading activation. However, the extent to which the N400 represents this process remains in question.

Collins and Loftus (1975) theorized that automatic semantic priming is due to how semantic information is organized in long-term memory (i.e., “semantic memory”). The authors suggested that semantic concepts in long-term memory are represented as “nodes.” The nodes are connected in semantic networks, with the distance of the connection between nodes reflecting the strength of the semantic association between them. For example, the semantic concepts for “apple” would have a shorter (i.e., stronger) connection to “orange” than “bacon” because the former two are both fruits. However, “apple” and “orange” would share a longer, weaker association with “peanut” since all three are included within the less specific semantic group of foods. Collins and Loftus (1975) theorized that encountering the word “orange” would be semantically primed by “apple” due to an automatic process called spreading activation. That is, when a node is activated, such as when presented in a semantic priming task, the activation of that word spreads to semantically related concepts. This activation is graded, such that the activation decays as it spreads over a longer distance. Thus, concepts that are weakly associated with an activated concept will be less activated (e.g., “peanut”) than those concepts that are more strongly associated with a concept (e.g., “apple”). This spreading activation to related concepts (e.g., spread from “orange” to “apple”) is thought to underlie behavioral and semantic priming effects. Thus, during lexical decision tasks, participants respond more quickly to words that are semantically related to a prime compared to non-related words. Spreading activation is also thought to explain ERP findings of semantic priming effects.

Similar to how participants respond more quickly to semantically primed words than non-primed words, individuals tend to generate less negative N400 amplitudes to targets that are preceded by a semantically related prime compared those preceded by an unrelated prime (e.g., Bentin et al., 1985; Kutas & Hillyard, 1989). However, these findings fail to establish the N400 as an automatic or controlled process (Posner & Synder, 1975) because both types of processing can be used if the probe is consciously perceived. Rather, the true test of the functional significance of the N400 as an automatic (lexical) or controlled (post-lexical process) comes from studies in which the prime is not consciously perceived, such as in masked semantic priming studies, attentional blink paradigms, and dichotic listening tasks. If N400 effects are generated to targets when the primes were not consciously perceived then it is thought to reflect an automatic semantic priming process. However, if the N400 effect is only generated to targets that were consciously perceived, then it is thought to represent a controlled priming process. If the N400 effect is observed under both conscious and unconscious perception, than the N400 may reflect both automatic and controlled semantic priming. The majority of this work comes from masked priming studies.

Masked priming studies are similar to typical semantic priming studies: Participants perform a lexical decision task to visually presented words. Words that are preceded by a semantically related prime are responded to more quickly than unrelated targets. However, in masked priming studies, the ability to use controlled processing is thought to be eliminated. This is because the primes are presented at a very short duration (e.g., < 40 ms) such that they are not presumed to be consciously perceived. That is, if primes are presented at longer stimulus onset asynchronies (SOAs), then the semantic priming process is an aggregate of automatic and controlled processing. Stimulus onset asynchrony (SOA) refers to the time between the temporal onsets of the prime and target.

However, when the prime is presented at a short SOA, individuals may only use automatic processing streams to process the stimuli.

Referring to the dual-process model (Posner & Snyder, 1975), the Brown and Hagoort (1993) sought to examine if the N400 represented an automatic process of lexical access or a controlled, higher-order process of semantic integration. To test this, the authors used the masked semantic priming paradigm. Pairs of words were visually presented in which the second word ("target") was preceded by a semantically related or unrelated word. For half of the subjects, the prime was "masked" and presented for only 40 ms. According to the Collins and Loftus (1975) model, semantically related "nodes" are linked in the mental lexicon. When a word is presented, the node for that word is activated and the activation spreads to related nodes. This is an automatic process. Based on this theory, targets that are "primed" by related words should be accessed faster both in masked and unmasked conditions. If the N400 represents this lexical access process, the N400 amplitude should also decrease to the primed target. In fact, the researchers found a less negative N400 to the unmasked primed targets, but not the masked primed targets. They concluded that the N400 does not represent an automatic process of lexical access but, rather, a controlled, post-lexical semantic integration process. When they used this procedure in a simple reaction time experiment with a different group of subjects, reaction times were quicker to targets that followed masked and unmasked semantically related primes. Taken together, these findings suggested that automatic spreading activation of primes to targets does occur but that the N400 does not reflect this process. However, as noted below, other researchers found opposite effects under different experimental conditions.

For example, Deacon and colleagues (2000) employed a masked semantic priming task with a within-subjects design and shorter prime duration. Whereas Brown and Hagoort (1993) held the prime duration constant at 40 ms, Deacon and colleagues (2000) determined each participant's prime recognition threshold (13-27 ms) prior to the ERP recording. Deacon and colleagues' (2000) results conflict with Brown and Hagoort (1993): N400 amplitudes were less negative to targets that followed semantically-related masked and unmasked primes compared to unrelated primes. Deacon and colleagues (2000) concluded that the N400 must represent an automatic process because of the semantic priming effects both in the unmasked and masked conditions. Findings from Grossi (2006) corroborated this conclusion.

Grossi (2006) presented prime words for 50 ms followed by a target word that was either unrelated to the prime, high related (80% association) or low related (20% association). Participants' behavioral accuracy to identify primes as words or nonwords was not statistically different, suggesting that the manipulation was successful. Similar to Deacon and colleagues (2000), Grossi (2006) reported a N400 effect (more negative N400 amplitudes) to unrelated targets compared to both high and low related targets. The author concluded that the N400 reflected an automatic, pre-integration process, such as lexical access, because the prime was masked and not consciously perceived. However, this failed to explain why N400 effects were not observed to targets when masked primes were presented at 40 ms (Brown & Hagoort, 1993).

Some authors suggest that Brown and Hagoort (1993) did not find semantic priming effects because of a long stimulus onset asynchrony (500 ms) that diminished masked priming effects (Kiefer, 2002). Said otherwise, automatic semantic priming may occur between prime and target (e.g., Deacon et al., 2000; Grossi, 2006) but only within a minimal time window after the prime (less than 500 ms). Kiefer (2002) obtained evidence for behavioral (i.e., reaction time) and N400 semantic priming (smaller N400 amplitudes to semantically primed targets) both in unmasked and masked conditions at an SOA of 67 ms. This finding replicated the author's prior work (Kiefer & Spitzer, 2000). However, at a longer SOA of 200 ms, semantic priming effects on the N400 were only found to targets following the unmasked primes (Kiefer & Spitzer, 2000). Taken together, findings from Kiefer (2002) and Kiefer and Spitzer (2000) provided direct evidence that automatic semantic priming effects on the N400 were likely absent in Brown and Hagoort (1993) because of the long SOA between prime and target (500 ms).

Kiefer (2002) also suggested that prior masked semantic priming effects on the N400 (e.g., Deacon et al., 2000; Grossi, 2006) could arise from "backwards priming" effects between the target and the masked prime. For example, the presentation of the target (e.g., "movie") could backwards prime the identification of the previously presented probe ("popcorn"), if the probe was initially partially identified ("_ _ pcor_"). To examine the influence of backwards priming, Kiefer (2002) conducted a follow-up experiment in which a "context word" was presented after the target word. If masked priming effects were due to backwards priming, then participants should be better able to identify the target word as a word (compared to a nonword) when it was followed by a semantically related context word. However, accuracy on this lexical decision task did not vary based on whether the context word was a word, nonword, repeated string of letters, nor when the participant was asked to make a semantic judgment about the relationship between the target and context word. The author concluded that the masked priming effects on the N400 were not due to controlled semantic processing (short SOA, 67 ms) nor backwards priming, suggesting that that N400 reflects an automatic process.

In addition to masked semantic priming studies, some researchers attempted to examine N400 effects when participants were prevented from engaging in semantic processing. For example, Chwilla and colleagues (1995) conducted an unmasked semantic priming task in which participants completed a lexical decision task or a physical discrimination task (i.e., identify if word was presented in uppercase or lowercase letters). The researchers' theorized that if the N400 were a reflection of automatic processing, then N400 priming effects would be present even when only surface-level characteristics of the prime-target pairs were processed (i.e., physical discrimination task). However, if the N400 effect depended on a controlled semantic integration processes, then the N400 priming effect would only be generated under situations of semantic processing demands (i.e., lexical decision task).

In support of their hypotheses, Chwilla and colleagues (1995) found that N400 amplitudes were more negative to non-semantically primed targets compared to primed targets in the lexical decision condition only, not the physical discrimination task. Based on this finding, the researchers concluded that the N400 could not represent an automatic process because semantic priming did not affect the N400 when semantic features of words were not processed. Furthermore, the authors' reported that reaction times to related and unrelated targets were the same in the physical discrimination task, suggesting that participants did not process semantic features of these words.

However, Orgs, Lange, Dombrowski, and Heil (2008) found that physical discrimination tasks did elicit N400 effects to environmental sounds that were semantically unrelated to visually presented word primes. In their study, participants completed two different blocks of a semantic priming task. In the first block, participants pressed a button in response to a sound if a word in blue ink preceded it but not if a word in red ink preceded it (Go/No-Go task). In the second block, participants were instructed to judge the semantic relatedness between prime word and target environmental sound and, again, did not make a behavioral response to No-Go trials. The researchers found N400 effects both in the semantic and physical conditions, albeit the N400 effects under physical discrimination task demands (automatic semantic processing) were statistically smaller than under the semantic priming task demands. The researchers concluded that automatic processing is sufficient to elicit N400 effects but that N400 effects were also a manifestation of controlled language processing.

Similar to other findings supporting the post-lexical account of the N400 (Brown & Hagoort, 1993; Chwilla et al., 1995), Bentin and colleagues (1995) did not find semantic priming effects on the N400 when participants did not attend to prime-target pairs. Participants in their study completed a dichotic listening task in which two different sets of words were presented simultaneously to each individual ear. The participants were instructed to only attend to words presented in one of the ears; however the prime and target pairs were always presented to opposite ears. Similar to Chwilla and colleagues (1995), the researchers hypothesized that if the N400 represents an automatic process then semantic priming effects on the N400 would appear even if the target word was presented to the unattended ear. Results showed that participants only generated an N400 effect when the target was presented in the attended ear. However, the SOA between prime and target in this study (1000 ms) may have been too long to pick up on the semantic priming effects to the unattended ear.

Importantly, Kiefer and Spitzer (2000) determined that a masked probe could induce semantic priming effects on a related target (unconscious semantic priming) if that target occurs within a 67 ms SOA of the probe. However, at a longer SOA of 200 ms, semantic priming effects were only found for the unmasked probes. Based on these findings with a 200 ms SOA, one can reasonably assume that the SOA of 1000 ms was too long to pick up on unconscious, automatic semantic priming effects in a dichotic listening task. Therefore, although Bentin et al. (1995) concluded otherwise, participants night have generated an N400 effect to targets following semantically related unattended primes, but this effect was not identified because of the long SOA (1000 ms).

Similar to findings from Orgs et al. (2008) but in contrast to other work (Chwilla, et al, 1995), Rolke and colleagues (2001) found significant semantic priming effects when the prime word occurred within an "attentional blink." As the researchers described, according to the attentional blink phenomenon (cf. Raymond, Shapiro, & Arnell, 1992), after the identification of a target, the ability to identify a subsequent target decreases if that target occurs within 200-500 ms following the first target. In this study, participants were instructed to identify visual words presented in white ink but ignore words presented in black ink. All words appeared on a gray background. Importantly, the prime occurred within the attentional blink (within 200-500 ms after a target), whereas the probe occurred outside the attentional blink. Accuracy in identifying the prime was significantly lower to primes compared to probes, indicating that the attentional blink manipulation was successful. However, participants still identified the primes on approximately 49% of trials. Therefore, the researchers split their analyses based on the N400 amplitudes to the probe words that followed attended and unattended primes. As expected, participants generated the typical N400 effect to probe words semantically associated with an attended prime. In addition, the N400 effect was reliable for probes that were semantically associated with an unattended prime, albeit the N400 effect was smaller for the probes following unattended primes compared to attended primes. Because controlled processing was minimized to unattended primes the researchers concluded that automatic processing was sufficient to elicit N400-semantic priming effects. Because these effects were smaller than attended semantic priming, they also concluded that the N400 might not reflect a completely automatic process but might also reflect some aspect controlled processing. Other similar research reported the same findings: N400 amplitudes to semantically primed probes occurring in an attentional blink were equivalent to those occurring outside the attentional blink (Luck et al., 1996). Taken together, findings from attentional blink studies further substantiate the point that the N400 effect is robust even when controlled processing is eliminated, thereby indicating that the N400 is sensitive to automatic spreading activation mechanisms.

Other research suggests that the N400 indexes both automatic and controlled processes. Holcomb, Reder, Misra, and Grainger (2005) suggested that automatic priming effects in prior masked priming studies might be due to conscious perception of primes being "leaked through." Therefore, based on participants' behavioral accuracy of identifying probes, the researchers analyzed differences in the N400 effect to semantically primed words based on this "prime visibility." As expected, they found evidence for semantic masked priming across prime presentation durations of 40 ms, 80 ms, and 120 ms. However, in the 40 ms condition, the amplitude of the N400 was negatively related to participants' identifications of the primes: participants who identified more primes generated more negative N400 effects to semantically primed targets, whereas participants who identified fewer primes had less negative N400 effects to these targets. This finding led the authors to conclude that the N400 effect may be sensitive to automatic pre-lexical processes in part but that this does not explain the fact that N400 effects were larger when consciously perceived (post-lexical process).

Based on the previously reviewed findings, both automatic and controlled semantic priming appear to be sufficient to influence the amplitude of the N400 ERP: Words preceded by semantically related primes elicit less negative N400 amplitudes compared to words preceded by unrelated primes. This is exemplified when the prime is consciously perceived (Brown & Hagoort, 1993; Deacon et al., 2000) and when conscious perception is eliminated (Deacon et al., 2000; Grossi, 2006; Keifer, 2000) or minimized (Rolke et al., 2001). Although some researchers failed to find evidence for automatic semantic priming on the N400 (e.g., Brown & Hagoort, 1993; Bentin et al., 1985; Chwilla et al., 1995), this may be due to long SOA durations between prime and target—automatic semantic priming may only occur within a limited time window following a prime (cf. Kiefer, 2002; Kiefer & Spitzer, 2000). On the other hand, at longer SOAs, controlled processing may occur and influence N400 amplitudes. Next, we turn our attention to how wider contexts, such as those from sentences and discourses, may prime the N400 effect.

1.3.1.1.2. Evidence from sentence and discourse comprehension tasks

As reviewed above, Kutas and Hillyard (1980) first discovered the N400 as an increased negative deflection between 300-600 ms following the presentation of anomalous sentence-final and sentence-intermediate words (Kutas & Hillyard, 1983; Kutas, Van Petten & Besson, 1988). Similarly, Kutas and Hillyard (1984) examined the N400 to words at the end of sentences that varied based on their level of clozure (e.g., expectancy) within the sentence (high, medium, low) and the level of constraint of the sentence (high, medium, low). In brief, clozure refers to the probability of a sample of participants to use a specific word to complete a sentence. Their results showed that the amplitude of the N400 was positively related to the clozure of the word within the sentence, leading the authors to propose that the N400 is related to expectancy of a word within a sentence context. In a follow-up analysis, the authors examined the N400 amplitudes to low cloze probability words in the high constraining contexts. They split these words into groups based on their relation to the word best completing that sentence. They found that N400s were less negative to the low cloze probability words semantically related to the best completion. This led the authors to conclude that the low-clozure words semantically related to the best completions were "primed" (Collins & Loftus, 1975) and thus had a less negative N400. Therefore, these authors concluded that the N400 might reflect, at least in part, an automatic semantic priming mechanism.

Kutas and Hillyard (1984) were the first to report that the amplitude of the N400 component to semantically coherent words was related to the cloze probability of that word, regardless of semantic constraints. Others replicated this finding (DeLong, Urbach, & Kutas, 2005). For example, when all critical words were semantically coherent, N400 amplitudes were still significantly more negative to low cloze probability words in high, medium and low constraining contexts compared to high cloze probability words in a high constraining context (Kutas & Hillyard, 1984). Importantly, there was no significant difference in N400 amplitude between the low cloze probability words depending on the level of semantic constraint in the sentence. However, words that were coherent, related, but unexpected elicited more negative N400s than the expected words, although the size of this amplitude was less negative than that to semantic anomalies. They concluded that varying levels of offline expectancy of coherent words can elicit the N400 effect (Kutas & Hillyard, 1980) and that the N400 was more likely an index of expectancy than contextual constraint. In sentence-contexts, unexpected words (based on offline cloze procedures) elicited more negative N400 amplitudes compared to expected words (Kutas & Hillyard, 1980; Laszlo & Federmeier, 2009). N400 effects have also been found when a context was provided by a larger discourse, such as a paragraph or story. However, researchers disagree on the extent to which N400 effects within discourses represent message-level effects or local word-by-word priming effects.

It is worth noting that some researchers suggest that N400 amplitudes are equal between low- and high-cloze probability words if they coherently complete a discourse passage. Van Berkum, Zwitserlood, Hagoort, and Brown (2003) reported that N400 amplitudes to coherent discourse-completing words were equal when elicited by words with a low cloze probability (mean = 1%) or high cloze probability. Importantly, this was only established when the discourse-completing word was coherent with the prior context—amplitudes to these coherent words were significantly smaller compared to low-cloze probability incoherent discourse-completing words that replicated the typical N400 effect (Kutas & Hillyard, 1980, 1983, 1984). This finding suggests that the cloze probability of a word may only influence the N400 amplitude to some extent, such as when the word has a low-cloze probability and is incoherent within the context.

For example, Otten and Van Berkum (2007) designed two-sentence stories in which a word within the second sentence was expected (high cloze > 0.50) or unexpected (low cloze < 0.30). The message-level representation of the sentence was also manipulated such that the high- and low cloze probability words were placed in stories that provided a coherent message or a message-level anomaly. Importantly, the same semantically relevant words occurred both in coherent and anomalous stories. The researchers examined the N400 effects to semantically related words within a coherent discourse, semantically related words within an anomalous discourse, unrelated words within a coherent discourse, or unrelated words within an anomalous discourse. If semantic relations between words independent of message-level information accounted for N400 effects, then coherent message-level and message-level anomaly conditions would both generate N400 effects. In contrast, the authors found that the N400 effect was only generated to the semantically primed target word in the coherent message-level condition, not when words were semantically related in an anomalous discourse, suggesting that the N400 reflects message-level constraints, at least in part. Interestingly, the low cloze probability words in the message-level anomaly condition elicited more negative N400 amplitudes at left anterior electrode channels (referenced to average of linked mastoids) compared to the message-level congruent condition. This suggests that a different neural mechanism may subserve the effect of message-level priming.

Van Berkum, Hagoort, and Brown (1999) also theorized that the N400 was sensitive both to local (sentence-level) and global (discourse-level) semantic fit. Participants in their study read discourses of three sentences in length. The researchers compared the N400 ERP elicited to semantically coherent words that either fit or did not fit with the preceding context: N400 amplitudes were larger to the discourse anomalies compared to the discourse coherent words. The authors also examined the third sentence of each paragraph in isolation in a different sample. Importantly, what was once a discourse anomaly was not anomalous when taken out of the context. They found that N400 amplitudes were significantly less negative to the discourse anomaly when it was taken out of the context (isolated sentence) compared to when it was a part of the discourse. This led the authors to conclude that the N400 was sensitive to discourse-level contextual fit because they compared the exact words but with or without the prior context. In addition, the more negative N400 to the discourse anomaly within the context was similar to a semantically anomalous condition within an isolated sentence. In fact, the N400s were equivalent in amplitude and topography, albeit with a significantly earlier onset for the discourse anomaly N400. Therefore, the authors concluded that the processes supporting semantic integration into sentence- and discourse-level contexts overlapped and shared similar underlying neural generators.

Federmeier and Kutas (1999) suggested that both automatic priming and message-level priming might differentially influence the N400. ERPs were recorded while participants read two-sentence long stories, such as “They wanted to make the hotel look more like a tropical resort. So along the driveway they planted a row of—.” The last word of the story was either expected (--palms), an unexpected within-category violation (--

pines), or an unexpected between-category violation (--tulips). Importantly, both the within- and between-category violations were coherent, unexpected sentence endings. The authors found that N400 amplitudes were significantly more negative to the between-category violation compared to the expected word. However, the N400 amplitude to the within-category violation was more negative than the expected word ending but less negative than the between-category violation. Based on the smaller N400 to the expected words, the authors suggested that the N400 was sensitive to message-level information (i.e., "palms" is a better fit than "tulips"). In addition, the authors also stated that the N400 reflected some level of long-term memory processing. They concluded this because the words with shared semantic feature overlap (i.e. within-category violation) generated an N400 that was smaller than the between-category violation even though the within- and between-category violations had the same cloze-probability. In their summary, they suggested that during sentence processing a specific context generates a context-dependent list of semantic features that are expected to appear within the sentence. Therefore, N400 amplitudes reflected the level of match between the presented word and the expected word. In short, the researchers suggested that context serves to influence semantic feature activation in long-term memory.

It is worth noting that although N400 effects may be sensitive to sentence-level (Kutas & Hillyard, 1980) and discourse-level contextual priming (Van Berkum et al., 1999), these findings do not indicate that the N400 represents an integrative process per se. In other words, the largest context available (e.g., a word, sentence, or a story) primes the lexical-activation of newly presented word (cf. Van Berkum, 2004, 2009). However, this does not suggest that the process by which that word is consciously integrated into the prior discourse is a controlled process. Rather, these findings may simply suggest that automatic spreading activation to target words may come from wider conceptual contexts (e.g., a story) and not just single words. Therefore, findings from Federmeier and Kutas (1999) suggest that the mental representation of a wider context (e.g., a tropical resort) pre-activates the access of contextually relevant words. The words that are pre-activated not only depend on semantic relations with prior words but also on the contextual representation of the story (Van Berkum, 2004, 2009).

Based on their findings (Federmeier & Kutas, 1999), Kutas and Federmeier (2000) theorized that the N400 represents the neural activity supporting the ability to retrieve lexical information from long-term semantic memory. In summary, they proposed that the language comprehension system pre-actives semantic memory features of expected words (i.e., predicts) during language processing. Importantly, they posited that the N400 also could not represent the level of a word's plausibility within a sentence (i.e., semantic integration hypothesis) because the N400 amplitude to within-category violations was less negative than that of between-category violations, even though the two were equally implausible words to finish the two-sentence discourses (Federmeier & Kutas, 1999). In addition, the authors reviewed findings of more negative N400 amplitudes that occurred as soon as 200 ms after the presentation of incongruent sentence-final words—too quickly for people to identify the actual word. However, when participants discriminated between congruent and incongruent sentence-final endings that shared their first phoneme, the N400 effect to the incongruent sentence-final ending was delayed. Taken together, Kutas and Federmeier (2000) concluded that their work and that of Federmeier and Kutas (1999) suggests that the language comprehension system uses any available contextual information to pre-active semantic features of future words even before they are presented. Based on this theory, the researchers hypothesized that words sharing semantic features within a context were more easily accessed and led to a reduction in N400 amplitude, whereas words not (or less) pre-activated required more effortful semantic processing as evidenced by larger N400 amplitudes.

Nieuwland and Van Berkum's (2006) findings support this theory. In their study, participants listened to sentences either about an animate object or inanimate object performing human-like behaviors. When participants first encountered the inanimate object doing human-like behaviors (e.g., a yacht going to a psychotherapist), they generated more negative N400 amplitudes. However, when the fictitious discourse of the story became clear, participants failed to generate N400 effects to animacy violations. In addition, N400 amplitudes were larger to expected predicates (e.g., peanut being salted) compared to when the context of the story was about a peanut performing an animacy violation (e.g., a peanut falling in love). Therefore, the N400 fluctuated in relation to its relevancy within the discourse model, not to our world knowledge of what inanimate objects can and cannot do. This led the authors to suggest that global context effects can override local semantic violations as represented by increased N400 ERPs to context-violations that were semantically coherent within the local sentence.

The authors proposed that the N400 represents the quick integration of words into the mental model of a discourse. However, these results are not necessarily at odds with the lexical-semantic retrieval account of the N400 (Kutas & Federmeier, 2000; Brouwer et al., 2012). If we assume that spreading activation (Collins & Loftus, 1975) influences the pre-activation of semantically related words in a discourse, we can explain why N400 amplitudes would be smaller to contextually relevant predicates that violate local semantic anomalies (e.g., about a peanut dancing). If so, this would suggest that contextual information eases lexical-semantic retrieval regardless of the real-word possibility of the context.

Further evidence to support the automatic account of the N400 comes from Federmeier, Wlotka, De Ochoa, and Kutas (2007). In this study, sentences were either weakly or strongly constraining, and sentence final endings were either expected (high cloze probability) or unexpected (low cloze probability). The unexpected words were chosen from a cloze probability task in which participants completed sentences with the “best completion” but also provided two alternative completions. Importantly, the unexpected words in the weak and high-constrained sentences were matched on cloze probability so the researchers could examine the independent effects of sentential constraint on the N400. In contrast to Federmeier and Kutas (1999), Federmeier and colleagues (2007) chose unexpected endings that did not share semantic features with the expected endings. Assuming the theory of spreading activation (Collins & Loftus, 1975), we would not expect reduced N400 amplitudes to coherent, unexpected endings in strongly constraining sentences because the unexpected endings did not share semantic features within the sentence. This is exactly what the researchers found: N400 amplitudes to unexpected endings were the same in the strongly and weakly constraining sentences. Based on this finding, we can conclude that the strong constraining sentence did not “prime” the coherent but unexpected word because this word did not share lexical features with the sentence.

The lexical-semantic retrieval account of the N400 may explain null N400 effects to semantically congruent but contextually anomalous sentence endings. For example, Fischler and colleagues (1983) conducted a study in which participants read sentences in which the object was either true/false, and the verb of the sentence made the sentence either positive/negative. They used four types of sentences: (a) true-affirmative (e.g., "A robin is a bird"), (b) true-negative ("A robin is not a truck"), (c) false-affirmative (e.g., "A robin is a truck"), and (d) false-negative ("A robin is not a bird"). The researchers focused their analyses on the ERPs recorded to the final word in each sentence. As expected, participants generated significantly more negative N400 amplitudes to sentence-final words in the false-affirmative sentences (e.g., "A robin is a truck") than those in the true-affirmative sentences (e.g., "A robin is a bird"). Surprisingly, the N400 amplitudes were also more negative to true-negative ("A robin is not a truck") statements compared to false-negative statements ("A robin is not a bird").

The semantic integration theory of the N400 (e.g., Hagoort, 2003, 2008) cannot explain these results. According to this hypothesis, the false-negative statement would elicit a more negative N400 because the final word ("--bird") is semantically incongruent with the local sentence ("A robin is not a—"). However, their findings showed that the semantically congruent, true-negative sentence produced the more negative N400—"A robin is not a truck," not the semantically incongruent statement—"A robin is not a bird." In contrast, the lexical-semantic retrieval account of the N400 can explain this finding (Kutas & Federmeier, 2000; Brouwer et al., 2012). According to this account, the word "robin" pre-activates the semantic features of "bird," and thus the latter is more easily integrated into the context and produces a less negative N400 even if it is an anomalous ending to the sentence. However, the word “truck” in the true-negative statement is not semantically associated with “robin.” Therefore, the semantic features of “truck” are not pre-activated in semantic memory and thus additional processing is required to retrieve the meaning of this word. Therefore, the N400 was not reflecting the integration of the sentence-final word within a sentential context but rather reflected the ease in which semantic information of words (i.e., their “meaning”) was accessed from long-term memory.

The lexical-semantic retrieval account (e.g., Kutas & Federmeier, 2000) of the N400 also can explain findings from Nieuwland and Van Berkum (2005). As reviewed at the beginning of this section, Nieuwland and Van Berkum (2005) failed to find an N400 effect to words anomalous with the message-level representation of a discourse but semantically related to the content words within the discourse (“semantic illusion phenomenon”). Findings from Nieuwland and Van Berkum (2005) converged with Myers and O’Brien’s (1998) theory of text comprehension. That is, the language processor follows a “dumb” process in which semantic elements of presented words “resonate” to semantically similar words in long-term memory and brought into working memory regardless of their fit with the message-level representation. This theory would explain why N400 effects were “skipped over” to words semantically related to the theme of a discourse but incoherent with the message-level representation of the discourse. Rather, the researchers found that the P600 ERP was sensitive to the message-level violation, representing the “checking” process of elements in working memory to the discourse representation (Myers & O’Brien, 1998).

St. George, Mannes, and Hoffman (1994) conducted a seminal study on the extent to which global context influences the N400. Participants in this study read four individual paragraphs presented visually one word at a time. Importantly, the context of these paragraphs was ambiguous unless the paragraph was preceded by a title (Bransford & Johnson, 1972; Dooling & Lachman, 1971). Prior research using these paragraphs indicated that, without a title, the ability to recall words and themes from the paragraphs was significantly degraded (Bransford & Johnson, 1972; Dooling & Lachman, 1971). In this study (St. George et al., 1994), participants read the paragraphs for comprehension and did not make a behavioral response. The researchers focused their analyses on the changes in N400 amplitude between participants who read the paragraphs with titles ("Titled" condition) and without titles ("Untitled" condition). N400 amplitudes, averaged across each word in the paragraphs, were significantly more negative for participants in the Untitled condition compared to the Titled condition. Based on this finding, the authors concluded that the N400 likely reflected an integrative process by which a word is connected to the global context of the story.

However, the Retrieval-Integration account of the N400 (Brouwer et al., 2012; Brouwer & Hoeks, 2013) and other theories (e.g., Kutas & Federmeier, 2000) suggest that the context provided by the title facilitates lexical-semantic retrieval of words. The plotted ERP waveforms in research by St. George and colleagues (1994) suggest that a late-positive potential followed the N400 effect between Titled and Untitled conditions: P600 amplitudes are qualitatively larger in the Titled condition compared to the Untitled condition. If significant, this finding would coincide with Brouwer and colleagues (2012) theory that N400-P600 effects occur in biphasic cycles. Specifically, the presentation of a title provides a context by which lexical information of words can be automatically accessed (N400). Secondly, without a context by which to interpret a discourse, integrating each word into the mental model is more difficult (lower P600s in Untitled condition) because there is no stable mental model by which to update.

In summary, burgeoning research suggests that the N400 represents an automatic process of lexical access or lexical-semantic retrieval. The N400 is positively related to its lexico-semantic fit with a word prime (Deacon et al., 2000; Grossi, 2006; Kiefer & Spitzer, 2000; Rolke et al., 2001), sentential context (Kutas & Hillyard, 1980, 1984), and discourse context (St. George et al., 1994). An automatic process of spreading activation (Collins & Loftus, 1975) is thought to account for N400 effects. That is, a prior context primes/pre-activates words that are semantically related to prior information in a sentence or discourse. Therefore, contextual information undoubtedly influences the N400 amplitude, but the N400 does not reflect contextual integration processes per se. Rather, other research suggests that the P600 component represents such combinatorial semantic processing. Such literature is reviewed in the following section.

The purpose of the present dissertation was to examine biphasic N400-P600 ERPs to coherent discourse in which the contexts of discourses were ambiguous without a title. However, when the contexts become clear (through a contextually relevant word), the P600, but not N400, should increase, thus representing the updating of one's mental model and identification of the theme. If supported, these findings would provide direct evidence to support the Retrieval-Integration theory (Brouwer et al., 2012).

1.3.2. The P600 ERP

Historically, researchers theorized that the P600 demarks the neural processing of syntactic errors and/or syntactic complexity (Friederici, Gunter, Hahne, & Mauth, 2004; Hagoort, Brown, & Groothusen, 1993; Kotz, Frisch, Von Cramon, & Friderici, 2003; Osterhout & Holcomb, 1992). However, other research suggests that the P600 represents discourse context updating or the change in the mental representation of the context of a language sample (for review see Brouwer et al., 2012, Brouwer & Hoeks, 2013).

1.3.2.1. The functional significance of the P600 ERP

Burkhardt (2006) identified that the amplitude of the P600, but not N400, was larger to words that introduced novel information within a discourse context. In a two sentence context, the researcher manipulated the extent to which a referent in the second sentence was associated with the prior sentence: (a) given (i.e., the same), (b) inferred, or (c) new. New referents and inferred referents generated larger P600 amplitudes compared to the given referent. Importantly, the N400 amplitudes to the given and inferred referents were both less negative than the new referent, with the N400 to inferred new referents being significantly more negative than the given referent. We can interpret these findings as they relate to the Federmeier and Kutas (1999) study. To briefly review, within-category violations within a discourse generated less negative N400 amplitudes than between-category violations. This is likely a result of the semantic-relatedness of the within-category violation to an expected exemplar in long-term memory (Federmeier & Kutas, 1999). Similarly, N400s to new referents in Burkhardt (2006) may relate, at least in part, to the easier retrieval of lexical information to these words compared to new referents. We may presume that this processing was needed to make an inference (due to the relation with the given information) compared to new information that bears no semantic resemblance to the given information. Although Burkhardt (2006) failed to take a strong opinion on the functional significance of the N400, she proposed that the P600 reflected the ease with which a word was integrated within a context.

Other research found a similar N400 effect of inferencing (St. George, Mannes, & Hoffman, 1997) but did not report on findings in the P600 latency range. However, N400 amplitudes in the inferencing condition were significantly reduced compared to a word-priming condition that contained the same content words as the inferencing condition but did not require forming an inference. This finding does not support the lexical-semantic retrieval theory of the N400 because the content words that would "prime" the inferencing were identical both in inferencing and word-priming conditions. However, St. George and colleagues (1997) averaged ERPs across every word in the sentence requiring inferencing rather than isolating the word in which the inference had to be made (Burkhardt, 2006). These differences in stimulus construction and ERP recording methodology may explain why St. George and colleagues (1997) did not report on P600 effects. Further research indicated that the P600 ERP might reflect inferential processing.

Burkhardt (2007) established that the amplitude of the P600, not the N400, positively relates to the strength of inferred association between a target word and cue during discourse processing. The first sentences of two-sentence discourses were manipulated such that the association between a target and a cue (in the second sentence) either was evident, probable, or inducible. Specifically, the discourse entity (e.g., "pistol" when a prior sentence included the word "shot") in the evident condition did not elicit an enhanced P600. However, the P600 was larger when the relation between the target and cue had to be inferred both in the probable and inducible contexts (e.g., “pistol” when a prior sentence included the phrases “killed” or “found dead”). Importantly, the amplitude of the N400 did not change based on the strength of the inferential association between cue and target. Based on these findings, the author concluded that the enhanced P600 reflects an additional processing cost necessary to establish a new discourse referent or re-organize the mental model of the discourse. Because these sentences were syntactically correct, these findings fail to support the functional significance of the P600 as a marker of syntactic repair only (cf. Hagoort, 2003, 2008). Furthermore, these findings converged with the assumption that the N400 reflected lexical-semantic processing (Kutas & Federmeier, 2000). Specifically, the three different types of targets (e.g., “shot,” “killed,” and “found dead”) may equally prime (at least statistically) the semantic activation of the cue (“pistol”), as represented by the null N400 effects across contexts.

Findings from others suggest that a late positive component (LPC) similar to the P600 reflects increased processing necessary to update a context of a discourse when establishing a new discourse referent. Kaan, Dallas, and Barkley (2007) investigated ERPs to discourse quantifiers in which a word phrase (e.g., “six flowers”) referred to a subset of a previously presented amount of objects (e.g., “eight flowers”) or a larger amount of that object (e.g., “four flowers”). Behavioral evidence from an offline study of sentence completion suggested that, when the first word of a second sentence referred to a subset of a number of objects from a prior sentence, participants were more likely to continue the second sentence with the same referent from the first sentence. However, if the quantifier in the second sentence was larger than the number of objects from the first sentence, they were more likely to continue the second sentence with a unique discourse referent. When a different group of subjects read these same discourses during EEG recording, the authors found a larger LPC (900-1500 ms post-quantifier onset) when the quantifier in the second sentence was a larger number than that presented in the first sentence as compared to when the quantifier was a smaller number (subset) of the prior sentence.

The authors concluded that this augmented LPC represented increased processing costs for establishing a new discourse referent because participants assumed that a larger number indicated the presentation of a new referent rather than a subset of the previously presented referent. The extent to which this LPC differs from the P600 is unclear. However, the authors did find a larger positive deflection ERP within the P600 time-window (500-700 ms) for individuals who were worse at answering comprehension questions about the discourses. A larger P600 in a similar time window was observed to words leading to inference making (Burkhardt, 2007). The authors suggested that this P600 might reflect a revision of the discourse model for those who are less accurate comprehenders. Therefore, the authors rejected the purely syntactic account of the P600 (Hagoort, 2003, 2008; Hagoort et al., 1993).

Nieuwland and Van Berkum (2005) found that unexpected words within a predictive discourse context generated enhanced N400 and P600 amplitudes compared to expected nouns. However, the unexpected words in a non-predictive context generated an N400 but no P600. Importantly, the content words were the same in both contexts. Because the N400 effect was present both in predictive and non-predictive contexts, the authors concluded that N400 amplitudes likely fluctuate due to automatic priming processes. However, they concluded that the P600 might index the occurrence of an unexpected message-level event. Because the P600, but not N400, fluctuated based on the incongruency with message-level information, we can presume that the former represents a post-lexical integrative process.

Similarly, Kim and Osterhout (2005) reported that verbs generated P600, but not N400, effects when they were placed in a semantically anomalous but syntactically correct sentence (e.g., "The hearty meal was devouring"). The researchers suggested that participants incorrectly established that the syntactically correct sentence was anomalous because the inanimate object ("meal") was the agent. Because there was no N400 effect, but a P600 effect, to "devouring," the researchers concluded that semantics might have overridden syntax and led to an incorrect analysis of the sentence as syntactically incorrect. In this sentence, the verb "devouring" was semantically related to the noun "meal." However, when the verb was semantically unrelated within the sentence (e.g., The dusty table tops were devouring), participants generated an N400 effect but no P600 effect. Similar to their initial finding, the authors suggested that the lack of semantic relation between "dusty" and "devouring" led participants to make the correct syntactic judgment of the sentence.

However, as reviewed in Brouwer et al. (2012), an alternative account can explain these findings. According to the Retrieval-Integration theory (Brouwer et al., 2012), the lack of N400 effect to "the hearty meal was devouring" results from the semantic relation between "meal" and "devouring." In addition, the P600 effect to the final verb in this sentence represents participants' revision of their mental representation of the sentence, not the incorrect judgment of the sentence as syntactically incorrect. This may explain why N400 effects were observed when “devouring” was placed in a sentence about “dusty tabletops” but a P600 effect was observed when “devouring” was placed in a sentence about a semantically related agent (“hearty meal”). According to the Retrieval-Integration account (Brouwer et al., 2012), the N400 effects were generated to verbs not semantically related within the sentence because lexical-semantic retrieval of such words is not pre-activated by lexical associates. The P600, on the other hand, was generated when the mental model of the sentence needed to be restructured. However, the fact that there was no P600 effect observed to “devouring” within the sentence about the “dusty tabletops” seems to contradict the Retrieval-Integration theory. The reason for such null effect warrants further investigation.

Findings from Sanford, Leuthold, Bohan, and Sanford (2011) provide direct evidence for the of the MRC account of the P600 (Brouwer et al., 2012). The authors examined N400 effects to discourses varying in the extent to which a semantic anomaly fit within the context (good fit, poor fit). In the good fit context, anomalies were semantically associated with the global context of the discourse. In the poor fit condition, semantic anomalies were not associated with the context and were thus easier to identify. The authors based their analyses on the fit of the context (good, poor) and whether the participants identified the anomaly. In the good fit condition, there was no difference in N400 amplitudes between the identified and unidentified semantic anomalies. Instead, the identified anomalies elicited a larger P600 amplitude compared to the unidentified anomalies, similar to that of Nieuwland and Van Berkum (2005). In the poor fit condition, anomalous words elicited larger N400s and P600s than coherent words. This finding supports Brouwer and colleagues' (2012) MRC hypothesis and biphasic account of N400-P600 effects. That is, the N400 was not influenced by local semantic anomalies because of its association within the context. Said otherwise, the prior context primes the semantic anomaly, likely through spreading activation (Collins & Loftus, 1975), even if the anomaly does not fit the context. However, the post-lexical integration process that connects the anomaly with the discourse was reflected by the P600 component. This was verified by the fact that the P600 was only generated when the anomaly was detected. Because the P600, but not the N400, was influenced by the identification of the anomaly, we can assume that the N400 is insensitive to global-fit violations when the anomaly is associated with the context.

It is possible that a more frontally distributed positivity (FP600) than the centroparietal P600 (Brouwer et al., 2012) may index complexity of discourse processing and/or the resolution of ambiguity in discourse. Kaan and Swaab (2003) manipulated the syntactical complexity and grammaticality of sentences. In doing so, they isolated the processes of syntactic revision and syntactic complexity without revision. Specifically, syntactically complex sentences included two noun phrases whereas syntactically simple sentences contained only one noun phrase. Both the one phrase and two phrase conditions contained sentences that were grammatical (syntactically correct) and ungrammatical (syntactically incorrect). As expected, the researchers found the typical posterior P600 effect for the syntactically incorrect sentences. However, they found the syntactically correct sentences in the two-noun phrase condition generated a larger frontal P600 compared to the syntactically correct sentences in the one noun phrase condition. Because both of these sentence types were syntactically correct, the authors concluded that the more frontal P600 might index discourse complexity whereas the posterior P600 might underlie syntactic revisions.

Other post-lexical accounts of the P600 suggest that it represents a conflict monitoring processes. Van de Meerendonk, Indefrey, Chwilla, and Kolk (2011) found topographically similar P600 effects to syntactic and spelling errors. To misspelled words within a sentence, the P600 effect was larger to high-cloze probability words compared to low cloze probability words. This led the researchers to suggest that the P600 reflected the greater conflict posed by the misspelled word in the high cloze probability condition compared to the low cloze probability condition. The researchers employed the same task during fMRI imaging and found that participants generated greater activation in the left inferior frontal gyrus both to syntactic and spelling violations compared to their correct counterparts. Based on these fMRI findings and the P600 effects, the authors concluded that the left inferior frontal gyrus subserves conflict-monitoring processing that generated the P600 ERP.

Similarly, Kolk, Chwilla, Van Herten, & Oor (2003) suggested that the P600 indexed conflict monitoring processes. In their study, semantic reversal anomalies and syntactically unacceptable sentences generated larger P600 amplitudes than their coherent semantic and syntactic counterparts. Importantly, the timing and morphology of these P600 effects were similar, leading the authors to suggest that they reflect the same underlying process. In contrast to Brouwer and colleagues (2012) who theorized that the P600 reflects the updating of a mental representation of a discourse, Kolk et al. (2003) suggested that the P600 reflects the identification of a conflict in the language processing stream between what is predicted and what is presented. However, Kolk and colleagues

(2003) posited that the N400 represented an integrative process. This is at odds with contemporary theories of the functional significance of the N400 as a lexical process (e.g., Federmeier & Kutas, 1999; Kutas & Federmeier, 2000). For a conflict to be identified, a word must, presumably, first be integrated into its prior context. Therefore, if the N400 in fact represents a lexical process (Kutas & Federmeier, 2000) and not an integrative post-lexical process (Kolk et al., 2003), the P600 cannot represent a conflict monitoring process because it first needs to be integrated with the prior context. Therefore, it is likely that the P600 represents this integrative process (Brouwer et al., 2012; Brouwer & Hoeks, 2013).

1.3.2.2. The influence of working memory on the P600

As previously reviewed, theories of language comprehension suggest that working memory is critical for effective discourse processing. For example, Clark and Haviland (1974) suggested that the language processor integrates new information into the mental representation of a discourse based on a matching process. Said otherwise, incoming information that matches or relates to previous information ("Given") is more easily integrated and understood compared to new information that does not match the prior representation of a discourse ("New"). Researchers have presumed that this matching process relies on, at least to some extent, working memory (Clark & Haviland, 1974; Kintsch, 1988; Kintsch et al., 1999; McKoon et al., 1996; Myers & O'Brien, 1998; Ratcliff & McKoon, 1988): Top-down knowledge from memory is necessary to integrate bottom-up information (Kintsch, 2005). For example, the ability to easily integrate the semantic representation of the word "apple" with a prior discourse about one's fruit preferences undoubtedly relies on the ability to remember the prior information presented. Importantly, research using ERPs supports Clark and Haviland's (1974) Given-New hypothesis: Participants generated larger P600 amplitudes to New information within a discourse compared to Given information (Burkhardt, 2006, 2007). This suggests that extra processing costs are necessary to integrate New information into a discourse compared to Given information.

Little debate exists on the role of working memory during language processing. However, researchers have yet to establish the neural mechanisms underlying working memory's role during language comprehension. Previous research suggests that working memory capacity, such as that assessed by the Reading Span Task (Daneman & Carpenter, 1980), moderated N400 effects within a sentence context (St. George et al., 1997; Van Petten, Weckerly, McIssac, & Kutas, 1997). Specifically, individuals read sentences that were congruent or anomalous, and ERPs were recorded to two words within each sentence type that were either semantically associated or unassociated (Van Petten et al., 1997). Individuals with a low reading span did not generate the typical N400 effect for semantically unassociated words within a congruent sentence-context (Van Petten et al., 1997). However, researchers have yet to establish the extent to which working memory capacity relates to the amplitude of the P600 ERP.

As reviewed, research suggests that amplitudes of the P600 are sensitive to a plethora of manipulations, such as syntactic violations (e.g., Hagoort, 2008; Hagoort, et al., 1993; Osterhout & Holcomb, 1992), misspellings (Van de Meeredonk et al., 2011), inferential processing (Burkhardt, 2007), and the establishment of new referents in discourses (Burkhardt, 2006; Kaan et al., 2007). Consequently, a modular hypothesis of the process(es) that generate the P600 ERP is insufficient to account for all of these findings.

Brouwer and colleagues (2012) proposed that their theory of the functional significance of the P600 could account for the range of P600 effects throughout the literature. Specifically, they suggested that the P600 is a family of late-positivities that reflect the revision of a mental model of a discourse (Brouwer et al., 2012). In short, the P600 reflects the processing cost necessary to revise the meaning of a sentence/discourse depending on new input into the language processor. That is, if the P600 represents this integration-revision process that is influenced by working memory (Clark & Haviland, 1974; Kintsch, 1988; Kintsch et al., 1999; McKoon, et al., 1996; Myers & O'Brien, 1998; Ratcliff & McKoon, 1988), then working memory abilities should modulate P600 amplitudes during discourse processing. In the next section, the neural generators of N400 and P600 effects are reviewed as support for this hypothesis.

1.3.3. Different neural generators of the N400 and P600 ERPs

Brouwer and Hoeks (2013) proposed that the left middle posterior temporal gyrus (lpMTG) subserves the retrieval of lexical-semantic information of a word. Specifically, more effortful lexical-semantic retrieval processing by the lpMTG generates an increase in N400 amplitudes recorded across the scalp. This information is then communicated via a series of white matter tracts (e.g., arcuate fasciculus, superior longitudinal fasciculus, and/or uncinate fasciculus) to the left inferior frontal gyrus (lIFG). Activity in the lIFG generates large P600 amplitudes when it is difficult to integrate a word within a linguistic context. Although the lIFG (e.g., "Broca's Area"), and particularly the ventral lIFG

(Costafreda, Fu, Lee, Everitt, Brammer, & David, 2006), is known for its role in semantic processing (Poldrack, Wagner, Prull, Desmond, Glover, & Gabrieli, 1999), researchers proposed that it might also support rehearsal in verbal working memory (Smith & Jonides, 1998), semantic memory-retrieval (Cabeza & Nyberg, 2000), or semantic working memory (Wagner, Desmond, Demb, Glover, & Gabrieli, 1997).

Cabeza and Nyberg (2000) reviewed literature suggesting that both the lIFG and lMTG may be involved in semantic memory retrieval, suggesting that the neuroanatomical localization of lexical-semantic retrieval (lpMTG) and integration (lIFG) as put forth by Brouwer and Hoeks (2013) may not be completely dissociated. For example, patients with damage to the dorsolateral PFC (including the left IFG) generated equivalent N400 effects as controls when processing incongruent sentence endings (Swick, Kutas, & Knight, 1998). However, the participants with brain damage failed to generate a posterior positivity between 600-900 ms to the incongruent sentence endings (i.e., "Late Positive Complex," P600), whereas the control group participants generated a prominent positivity in this latency range. These findings suggest that the lIFG generates the P600, not the N400. This integration process relies, at least to some extent, on working memory (e.g., Clark & Haviland; Kintsch, 1988, 2005; Kintsch et al., 1999) that also involves the lIFG (see Cabeza & Nyberg, 2000 for review; Wagner et al., 1997). However, the conceptualization of this working memory process as a domain-general or domain-specific (e.g., semantic working memory) mechanism remains to be established. The N400, on the other hand, is likely generated in a large part of the left temporal lobe (MTG: Brouwer & Hoeks, 2013; MTG/STG: Van Petten & Luka, 2006) and reflects a lexical process in which the semantic information of words is retrieved from long-term memory.

In their review, Lau, Phillips, and Poeppel (2008) proposed a cortical network for the semantic processing that underlies the N400 ERP. Reviewing evidence from ERP, fMRI, magnetoencephalography (MEG), and intracranial recordings, the authors concluded that the likely source of the N400 is the posterior middle temporal cortex, including the middle temporal cortex, superior temporal sulcus, and inferior temporal cortex. Their conclusion was largely based on fMRI findings using semantic priming paradigms that all converge on activation in this area both at short and long SOAs and related evidence from MEG. Importantly, this suggests that the MTG facilitates lexical access when semantic primes are both consciously perceived (long SOA) and not consciously perceived (short SOA). The authors also reviewed evidence that posterior inferior frontal regions support more effortful, controlled lexical selection when a word is not easily accessed. This led the authors to suggest that the late positivity (i.e., P600) that follows the N400 (Nieuwland & Van Berkum, 2005; see Brouwer et al., 2012 for review) may be a manifestation of activity in the posterior IFG.

Using MEG, researchers compared the two primary theories of the functional significance of the N400 (Lau et al., 2009)—that is, the N400 as a representation of lexical access/retrieval (Kutas & Federmeier, 2000, 2011; Lau et al., 2008) or rather a representation of semantic integration processes (Hagoort, 2008). They theorized that if the N400 represented an integration process, then the N400 response should be qualitatively different when a word is integrated within the context of a sentence (more representations) than when it appears in isolation (less to integrate with). However, if the

N400 represented lexical access/retrieval, then the N400 should be qualitatively similar between the two tasks.

In support of their hypotheses, the authors found that the N400 effect (larger amplitude to incongruous words), although stronger in the sentence completion task compared to the semantic priming task, was the same in topography and latency in the sentence completion and priming tasks. Unexpectedly, the authors also found a "late-N400 positivity" (cf. Van Petten & Luka, 2006) between 600-900 ms to incongruent sentence final endings compared to congruent sentence final endings. Although the authors did not elaborate on the functional significance of this component, they suggested it might reflect a semantic integration process. Again, this study provided evidence from a spatially and temporally sensitive imaging method on the time course of semantic processing of sentences—specifically, that incongruous sentence endings produce biphasic N400-P600 effects reflecting lexical access and semantic integration processes, respectively (cf. Brouwer et al., 2012). Taken together, reviews of neuroimaging literature (e.g., Brouwer & Hoeks, 2013; Lau et al., 2008, 2009; Van Petten & Luka, 2006) and lesion studies (e.g., Swick et al., 1998) provided further evidence that lexical access (automatic) and semantic integration (controlled) processes occur from activity in distinct neuroanatomical regions that might be differentially manifested as the N400 and P600 components.

In a follow-up to their ERP study (St. George et al., 1994), St. George, Kutas, Martinez and Sereno (1999) used fMRI to examine brain regions supporting discourse processing. Using the same stimuli at that used in their previous research (St. George et al., 1994), participants who read ambiguous paragraphs without titles generated greater brain activity in the inferior temporal sulcus than those who read titled paragraphs. In addition, whereas the Untitled group generated increased right hemisphere activation in the middle temporal sulcus, the Titled group generated greater left-lateralized activity in the same region. Across the two groups, participants generated significant activation in the left inferior frontal gyrus (lIFG) while reading the paragraphs. These findings corroborate previously reviewed theories regarding the role of the middle temporal loci as the neural source of the N400 (e.g., Brouwer & Hoeks, 2013; Lau et al., 2008; Van Petten & Luka, 2006).

As reviewed, participants who read ambiguous paragraphs without titles generated more negative N400 amplitudes than those who read the paragraphs with titles (St. George et al., 1994). Because researchers theorize that the left-middle temporal regions are the neural sources of N400 effects (Lau et al., 2008; Van Petten & Luka, 2006), it can be presumed that participants in the Untitled condition generated more activation in middle temporal areas (St. George et al., 1999) which led to larger N400 amplitudes (St. George et al., 1994). The likely cause of this increased activation was the lack of context available to facilitate lexical-semantic retrieval of words (Brouwer & Hoeks, 2013). Untitled groups' greater activation of right middle temporal regions likely underscored their more effortful processing to retrieve words, whereas the greater left middle temporal activation might represent the Titled groups' efficient lexical-semantic retrieval. However, the lack of lIFG group differences is surprising given that others suggested that the lIFG supports contextual integrative processing (e.g., P600)—a process that should be more difficult to Untitled than Titled paragraphs. A small sample size per condition (n = 5) may explain this null effect.

In interim summary, researchers disagree on the functional significance of the N400 and P600 ERPs. In short, some researchers theorize that the N400 ERP represents a controlled process in which the conceptual representation of a word is integrated into a prior context (e.g., Brown & Hagoort, 1993). Others propose that automatic lexical-semantic memory retrieval indexes the N400 component (e.g., Brouwer et al., 2012; Kutas & Federmeier, 2000, 2011). The main support for this hypothesis comes from studies in which N400s to target words were smaller following semantically related primes even when the primes were not consciously recognized (e.g., Deacon et al., 2000; Grossi, 2006; Kiefer & Spitzer, 2000; Rolke et al., 2001). In addition, words that were anomalous within a sentence, but semantically related with the theme of a discourse, did not elicit N400 effects but, rather, P600 effects (Fischler et al., 1983; Nieuwland & Van Berkum, 2005; Sanford et al., 2011). In opposition to the historical account of the P600 as an indicator of syntactic processing (e.g., Hagoort, 2003, 2008), recent theories suggest that the P600 may also represent the revision of a mental representation of a discourse (e.g., Brouwer et al., 2012; Brouwer & Hoeks, 2013) or the establishment of a new discourse referent (e.g., Burkhardt, 2006).

The Retrieval-Integration theory explains bi-phasic N400-P600 effects as indicating automatic and controlled processing respectively (Brouwer et al., 2012). Although a context pre-activates lexical-semantic retrieval (as represented by the N400), the P600 rather reflects the process by which the conceptual representation of a word is integrated into the mental representation of a context (Brouwer et al., 2012). This theory corresponds to Posner and Snyder's (1975) dual-process model of language processing. Findings from neuroimaging and lesion research suggest that the neural sources of the

N400 and P600 are likely in the left middle temporal and left inferior frontal gyri respectively (Lau et al., 2008, 2009; Swick et al., 1998; Van Petten & Luka, 2006). Based on theories of text comprehension (Myers & O'Brien, 1998) and discourse processing (Kintsch, 1988, 2005) and the shared role of the lIFG as underlying the P600 and working memory (e.g., Brouwer & Hoeks, 2013; Cabeza & Nyberg, 2000; Smith & Jonides, 1998), working memory capacity may moderate the P600 ERP component.

CHAPTER 2

STUDY 1: THE DEVELOPMENT OF A NOVEL STIMULUS SET FOR INVESTIGATING DISCOURSE CONTEXT UPDATING IN AMBIGUOUS NARRATIVE DISCOURSES

2.1. General introduction

This chapter reviews the stimulus development and construction of a novel method to isolate the processes in which the mental representation of a narrative discourse is updated. The results from this study, provided in Chapter 3, established a stimulus set of 25 narrative discourses. In Chapter 4, changes in N400 and P600 amplitudes during comprehension of these narrative discourses are reviewed (Study 2).

Prior research suggests that the N400 ERP indexes an automatic, lexical access or retrieval process (Brouwer et al., 2012; Brouwer & Hoeks, 2013; Burkhardt, 2006, 2007; Deacon et al., 2000; Federmeier & Kutas, 1999; Fischler et al., 1983; Grossi, 2006; Kiefer, 2002; Kiefer & Spitzer, 2000; Kutas & Federmeier, 2000; Lau et al., 2008, 2009; Luck et al., 1996; Rolke et al., 2001). Burgeoning research also suggests that the P600 ERP represents a controlled, post-lexical process when a word is integrated into the existing context of a discourse (Brouwer et al., 2012; Brouwer & Hoeks, 2013; Burkhardt, 2006, 2007; Sanford et al., 2011). Although the functional significance of the P600 continues to be widely debated, some researchers suggest that it indexes the process by which the mental representation of a discourse is updated/revised (Brouwer et al., 2012; Brouwer & Hoeks, 2013). From this point forward, this process will be referred to as "discourse context updating."

The present study sought to establish the functional significances of the N400 and P600 ERPs during the resolution of contextual ambiguity in narrative discourses. Initially, it was critical to develop a set of discourses in which one updates the mental model of the existing context after reading a specific word. If the P600 represents discourse context updating, its amplitude, but not that of the N400, should be larger to the word in which the mental model of a discourse is revised (Brouwer et al., 2012; Burkhardt, 2006, 2007). In Chapter 3, the Methods for examining the functional significance of the N400 and P600 ERPs during the processing of these discourses are reviewed. First, we report on the development and construction of the stimulus set of short, narrative discourses used to elicit the N400 and P600 ERPs during ambiguous discourse processing.

2.2. Purpose

There were two goals of this study.

Goal 1: Develop a set of discourses in which only the final word ("Critical Word 3") is highly semantically related to the contexts of the discourses.

To accomplish this goal, 80 discourses were constructed in which the contexts were initially ambiguous unless a title appeared before the discourse. The first two sentences of these discourses were developed to contain words in the sentence-final positions ("Critical Word 1," "Critical Word 2") that were not related to the contexts of the discourses. The third sentence contained a highly contextually relevant word in the sentence-final position ("Critical Word 3").

Participants in this initial study performed a cloze probability task and their performance was evaluated using the "cloze procedure" analysis developed by Taylor (1953). Taylor (1953) defined the cloze procedure as, "A method of intercepting a message from a 'transmitter' (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to 'receivers' (readers or listeners) that their attempts to make patterns whole again potentially yield a considerable number of cloze units" (1953, p. 416). A cloze unit is, "Any single occurrence of a successful attempt to reproduce accurately a part deleted from a 'message' (any language product) by deciding, from the context that remains, what the missing parts should be" (1953, p. 416). In short, the cloze probability of any word within a passage was calculated as the probability of the number of respondents who completed the mutilated passage with a specific word out of the total number of respondents. The level of cloze probability of an omitted sentence-final word depends on the level of constraint of the sentence (Bloom & Fischler, 1980). That is, highly constraining sentences (e.g., "She could tell he was mad by the tone of his—") produce high cloze probabilities ("–voice"; 99%; Block & Baldwin, 2010). On the other hand, low-constraining sentences (e.g., "There is something grand about the—") produce lower cloze probabilities ("–opera"; 22%; Bloom & Fischler, 1980).

The cloze procedure was originally developed as an index of the readability of a text and is not limited to mutilated words in the sentence-final position only (Taylor, 1953). However, this procedure has been commonly used in psycholinguistic research to develop well-controlled stimulus sets to study language comprehension using various brain imaging methods (e.g., ERPs, fMRI) both in normative (e.g., Laszlo & Federmeier, 2009) and clinical populations such as Alzheimer's disease (e.g., Revonsuo, Portin, Juottonen, & Rinne, 1998) and schizophrenia (e.g., Nestor et al., 1997; Pinheiro et al., 2015). As a result of the necessity for controlled research designs in these studies, cloze procedures are most commonly used to evaluate the expectancy of words in sentence-final positions.

Bloom and Fischler (1980) were the first to use the cloze procedure to develop a normative set of sentences for experimental research. Their goal was to develop two sets of sentential contexts. The first set was designed to elicit a number of unique responses (varied uncertainty set), and a second highly constraining set was designed to elicit a primary response (low uncertainty set). Using the cloze procedure, Bloom and Fischler (1980) developed 329 unique sentences. As expected, the varied uncertainty set produced a larger number of unique responses per sentence: The highest cloze probability for the majority of sentences was between 0.20 and 0.79. However, the majority of highest cloze probabilities in the low-uncertainty context were between 0.50 and 1.0.

To update a mental model of a discourse, participants must integrate a word with the prior information from the discourse (Brouwer et al., 2012; Brouwer & Hoeks, 2013).

Goal 2: When the context of a discourse is ambiguous, identify the word(s) that, after reading, leads to participants' ability to identify the context.

2.3. Materials & Methods

Stimuli in the present study were 80 three-sentence long discourses. Discourses were initially developed using four primary criteria: (a) sentences were grammatically correct when completed with a missing word in the sentence-final position, (b) at least two words could coherently complete the sentence, (c) no clichés, and (d) no sentences were “leading” (Block & Baldwin, 2010; Bloom & Fischler, 1980). Leading sentences are phrases with words that frequently co-occur, such as “peanut butter and jelly.” All sentences within each discourse were developed to be low constraining when presented in isolation. Prior research suggested that 5- to 10-word sentence contexts are optimal for providing maximum contextual constraints during cloze procedure tasks (Aborn, Rubenstein, & Sterling, 1959). Hence, all sentence contexts ranged from 4 to 9 words (7.12 ± 1.07) with only three sentence contexts that were four words in length. The majority of sentences (87.91%) were between 6-8 words including the final word (Bloom & Fischler, 1980). Discourses were coherent and described a subject (e.g., “the woman”) engaging in some sort of normal daily event. The events regarded everyday life activities with which all undergraduate students were presumed to be familiar.

2.3.1. Participants

Participants were 241 undergraduate students (Male = 45, Female = 196; 81.32% female) enrolled in psychology courses at the University of Nebraska-Lincoln (UNL). Participants were excluded from participating if they were less than 17 years of age or non-native English speakers (e.g., Deacon, Dynowska, Ritter, & Grose-Fifer, 2004). Participants were only included in analyses if they completed every item and/or their responses indicated that they accurately followed instructions. This sample size per Condition (approximately 30) is comparable to prior research using cloze analyses to develop new stimulus sets of written language samples (Block & Baldwin, 2010; DeLong et al., 2005; Laszlo & Federmeier, 2009).

The demographic characteristics of the sample appear in Table 2.1. Participants ranged in age from 17 to 47 years (20.11 ± 2.91). The average participant had completed 14.47 years of education (Range: 12-19; SD = 1.30), and the majority were Caucasian (88.5%). Participants in each Wave (i.e., data collection session) did not differ significantly in age, $F(7, 232) = 1.70$, $p = .110$. Participants in Waves differed significantly in years of formal education completed, $F(7, 232) = 6.03$, $p < .001$. In brief, years of formal education completed by participants in Wave 1-Untitled condition (15.52 ± 1.21) and Wave 1-Titled (14.90 ± 1.26) were slightly greater than both conditions in all of the other Waves. Years of formal education completed by participants in Wave 4-Untitled (13.83 ± 1.10) and Wave 4-Titled (13.94 ± 1.27) were slightly smaller than the both conditions in the other Waves.

2.3.2. Recruitment

Potential participants were recruited through the UNL Psychology Department Subject Pool via the SONA system. For the current study, a brief description of the study was available to potential participants. Participants received 3.0 SONA credits (1.0 credit for each half hour of participation) for participating. If applicable, this SONA credit could be applied to students' course credit or extra credit as determined by their course instructor. Individual meetings with undergraduate psychology courses provided another mechanism by which undergraduates were invited to participate. These presentations

Table 2.1

Demographic Characteristics of Sample by Condition and Wave

Condition	N	Age (M ± SD)	Gender[1]	Education[2] (M±SD)
Untitled				
Wave 1	29	21.25 ± 2.47^	89.70	15.52 ± 1.21
Wave 2	31	19.84 ± 1.49	90.30	14.35 ± 1.1
Wave 3	30	19.47 ± 1.20	76.67	14.23 ± 1.14
Wave 4	29	19.10 ± 1.21	75.90	13.83 ± 1.10
Untitled Total	119	19.90 ± 1.82	83.20	14.48 ± 1.288
Titled				
Wave 1	29	20.48 ± 2.34	86.20	14.90 ± 1.26
Wave 2	28	19.79 ± 1.62	89.30	14.71 ± 1.46
Wave 3	33	20.67 ± 4.93	72.70	14.36 ± 1.14
Wave 4	32	20.25 ± 4.41	75.00	13.94 ± 1.27
Titled Total	122	20.31 ± 3.66	80.30	14.46 ± 1.32
Sample Total	241	20.11 ± 2.91	81.70	14.47 ± 1.30

Note. [1] % Female, [2] Years of formal education completed, ^Age missing for one subject.

included an overview of the study procedures and the process by which individuals could participate. No consenting procedures were performed at this time.

2.3.3. Procedure & data analysis

Participants completed all study procedures individually and online. All participants provided electronic informed consent before participating in the study procedures. Participants read the consent form and decided whether or not they wanted to participate. Participants were instructed to complete the study individually and in a quiet environment without distractions so they could perform to the best of their abilities. Because of this format, completion of the study was self-guided. Participants were instructed that the study would take approximately 60-90 minutes to complete.

This study was a 2 (Condition: Titled, Untitled) X 4 (Wave: 1, 2, 3, 4) between-subjects factorial design such that participants completed 1 of 4 Waves and only one condition (Titled, Untitled) within each Wave. Waves were completed in sequential order. For example, 58 participants complete Wave 1 (29 in Titled Condition; 29 in Untitled Condition); then the next group participants completed Wave 2. Participants only completed one Wave. In the following section, the sequential data collection procedures and analysis approach are reviewed. Figure 2.1 demonstrates an example of the sequential data collection procedures for the Untitled group. The expected word for each sentence is provided in parentheses.

2.3.3.1. Cloze analysis

Participants were assigned into one of eight groups (i.e., Titled or Untitled conditions in each of the four Waves) depending on when they volunteered to participate in the study. Participants in the first six groups completed a cloze procedure. Participants in the last two groups did not complete the cloze procedure, but only a title identification judgment (described in section 2.3.3.2.). The cloze judgment was always made on the sentence-final position in each discourse. The groups varied based on (a) whether the title was presented prior to the discourse (Title: Titled, Untitled) and (b) the length of the

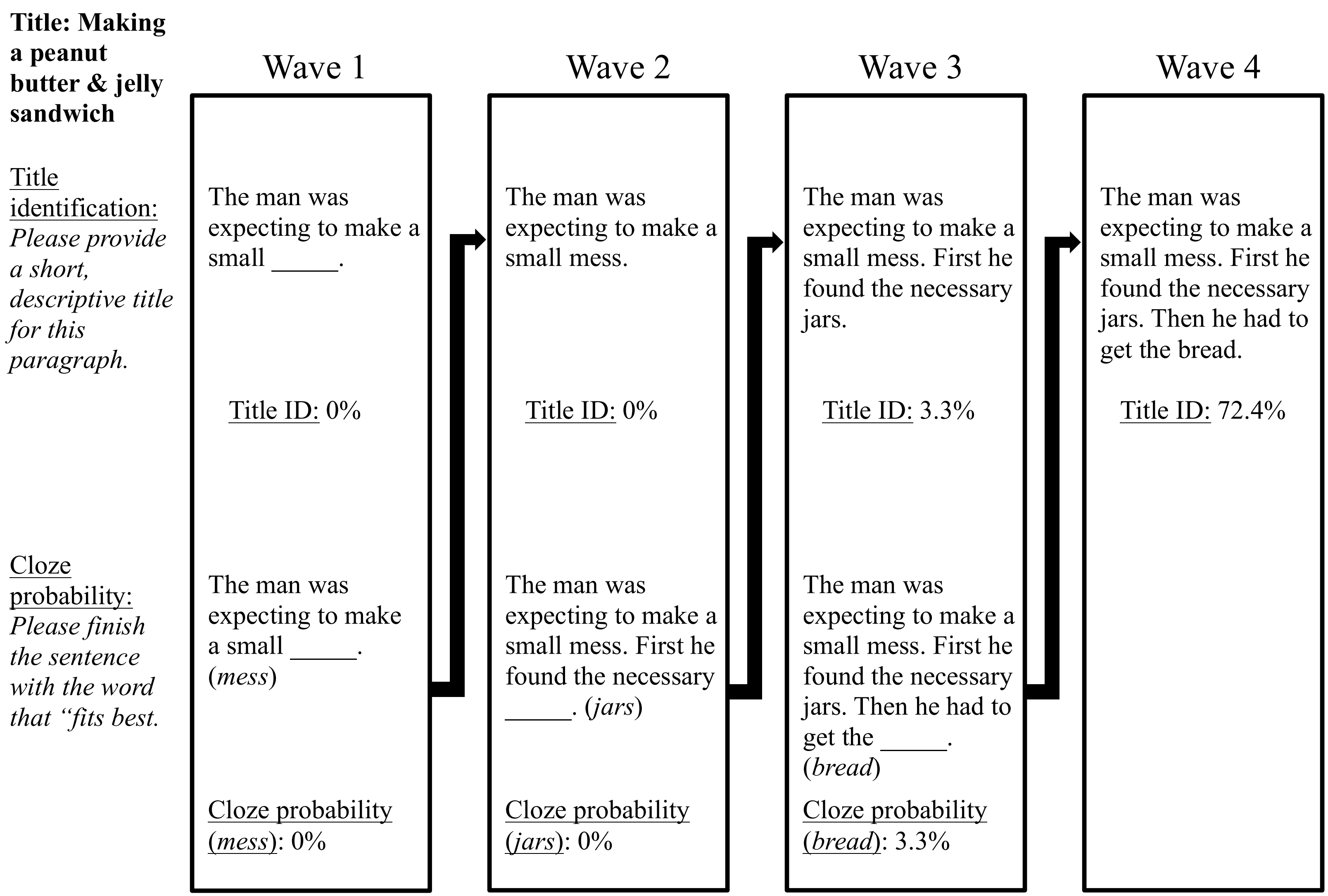

Figure 2.1. Sequential Data Collection Procedures for Study 1

discourse in which the cloze judgment was made on the last word in the discourse (Wave: Wave 1 = sentence 1, Wave 2 = sentences 1 & 2, Wave 3 = sentences 1, 2, & 3, Wave 4 = no cloze judgment). The groups participated in sequential order: Wave 1-Untitled, Wave 1-Titled, Wave 2-Unitled, Wave 2-Titled, Wave 3-Untitled, Wave 3-Titled, Wave 4-Untitled, Wave 4-Titled). Said otherwise, the Titled and Untitled conditions within Wave 1 were completed prior to the onset of Wave 2. This procedure was performed to establish that the sentence-final word chosen for sentence 1 was of low cloze-probability (Wave 1: Critical Word 1) before cloze judgments were made on Critical Word 2. For the same reasons, this established that the sentence-final word chosen for sentence 2 was of low cloze-probability (Critical Word 2) before cloze-judgments were made on Critical Word 3.

There is currently no established gold standard for determining “high” and “low” levels of cloze probability in sentences or discourses. For example, in a follow-up study to Bloom & Fischler (1980), Block & Baldwin (2010) examined the cloze probability of 498 highly constraining sentence contexts that included 100 of the low-uncertainty set (i.e., highly constraining) from Bloom and Fischler’s (1980) study. Block and Baldwin (2010) used strict cut-offs to establish low (0-33%), medium (34-66%), and high (67%-100%) cloze probability. However, other researchers have used lower cut-offs to establish relative levels of cloze probability.

For example, in an ERP investigation of language comprehension within a wider discourse, Van Berkum and colleagues (2003) developed coherent discourse passages with final words having an average cloze probability of 18% (which they considered high) and ranged from 0%-92%. Otten, Nieuwland, and Van Berkum (2007) classified high-cloze probability words in two-sentence discourses as those greater than 0.50 and low cloze as those less than 0.30, with a minimum difference between the two conditions of 0.25. Other researchers employed median-splits to dichotomize their stimuli into “low” and “high” constraints (e.g., Federmeier & Kutas, 1999). For example, Coulson, Urbach, and Kutas (2006) used a 40% cut-off.

For the purposes of this study, low cloze probability words were those used by less than 20% of respondents. High cloze probability was established as greater than or equal to 45%. This cut-off is similar to Coulson et al.’s (2006) 40% cut-off for high cloze probability and Block and Baldwin’s (2010) 34% cut-off for moderate-high cloze probability. This method is preferable than using a median-split (Federmeier & Kutas, 1999) or 30% as the high threshold for low-cloze probability words (Otten & Van Berkum, 2007) because it provides a greater contrast between “low” and “high” cloze probability words. Said otherwise, by using a 50% split, words with a 49% probability and a 51% probability would be placed into different groups (low v. high). However, cloze probabilities at the ceiling of the low cloze probability words (20%) and the floor of the high cloze probability words (45%) were substantially more disparate.

We will use the following discourse (“Making a Peanut Butter and Jelly Sandwich”) to demonstrate the hierarchical nature of the data collection process (see Figure 2.1). Participants in Wave 1 read the first sentence of the discourse with the final word omitted. Their task was to complete the sentence with the most appropriate “best guess” of the omitted word. The results from Wave 1 determined the sentence-final word for sentence 1. Based on participants’ responses, the research team chose a sentence-final word (Critical Word 1) that was coherent within the discourse and had a low cloze probability both in the Titled and Untitled conditions. This is because a high cloze probability word would likely facilitate participants' ability to identify the theme of the discourse. However, it was our intent to isolate this title identification process at the end of the third sentence (Critical Word 3) and not in the first or second sentences. The intended word to complete this sentence was "mess." No participant in the Untitled condition (cloze probability = 0.0%) completed the sentence with the word "mess," whereas one participant in the Titled condition (cloze probability = 3.4%) completed the sentence with this word. Therefore, we established low cloze probability both in the Titled and Untitled conditions for Wave 1. After establishing the cloze probability for the first sentence both in the Titled and Untitled conditions, data collection began for Wave 2.

In Wave 2, participants read the first and second sentences of the discourses with the sentence-final word omitted in the second sentence. Participants made their cloze judgment on this word. Based on participants' responses, the research team chose the sentence-final word to sentence 2 that was coherent with the discourse but had a low cloze probability both in Titled and Untitled conditions (Critical Word 2). The intended word to complete this sentence was "jars." No participant in the Untitled condition or the Titled conditions completed the sentence with the word "jars" (cloze probabilities = 0.0%). After establishing the cloze probability for the first and second sentences both in the Titled and Untitled conditions, data collection occurred for Wave 3

In Wave 3, participants read the first two complete sentences and made cloze judgments on the last word in the third sentence. Based on participants' responses the research team chose the word that had a high cloze probability in the Titled condition but a low cloze probability in the Untitled condition (Critical Word 3). The intended word to complete this sentence was “bread.” One participant in the Untitled condition completed the sentence with “bread,” (cloze probability = 3.3%), whereas 78.8% of participants in the Titled condition completed the sentence with this word. Therefore, we established low cloze probability of the critical word in sentences 1 and sentence 2 but a high cloze probability in sentence 3 in the Titled condition only.

During data processing, data were first screened to correct any misspellings and morphological errors (Block & Baldwin, 2010). For example, in the sentence “The man quickly cleared his—” the misspelled cloze unit “thoat” was corrected to “throat.” Pronouns were removed from entries that included a pronoun before the noun (e.g., “the throat” was revised to “throat”). Synonyms (e.g., car, automobile) were treated as unique words (Block & Baldwin, 2010). Singular and plural versions of the same nouns were treated as identical cloze units (e.g., car, cars). Compound words (e.g., paintbrush) that contained a portion of a unique cloze entry (e.g., brush) were treated as the same unit. Participant non-compliance was also determined based on the appropriateness of responses. Participants were excluded from analyses if they made errors (e.g., non-noun responses, included more than one word per response, incomprehensible word) on more than 10% of items. Otherwise, incorrect items (e.g., two words per one entry) were retained. Participants were not included in the final data set if they failed to complete all items.

2.3.3.2. Title identification analysis

In addition to making a cloze judgment for each discourse, participants made a title identification judgment. They were instructed to provide a descriptive title for each discourse. Figure 2.1 also demonstrates the title identification procedure for Waves 1-4. Note that Wave 4 included a title identification judgment for the full three sentences of each discourse, but no cloze judgment. Participants in Wave 1 first made their cloze judgment on sentence 1. Immediately following, they made their title identification for the discourse on the first sentence using their unique cloze unit. Participants in Wave 2 read the complete first sentence of the discourse and made their title identification based on this sentence. Participants then again saw the first sentence and also the second sentence with the missing sentence-final word. The participant made the cloze judgment on this sentence-final word of the second sentence. In other words, for each discourse in each Wave, participants made a title identification from the sentence(s) that were constructed from the cloze procedure from the previous Wave. Continuing with this pattern Wave 3 consisted of a title identification after reading the first two sentences then a clozure task of the last word in sentence 3.

The contexts of the discourses were provided in the Titled condition. However, for those in the Untitled condition, the contexts of the discourses should be ambiguous until the presentation of the high cloze probability word in sentence 3 (Wave 4). For example, one participant in the Untitled condition for Wave 3 provided the following title for the example from above: “Learning to craft.” For Wave 4, participants identified the title after reading the first three sentences (Critical Word 3 based on the clozure from Wave 3).

Objective scoring criteria were established for determining the accuracy of participants' title identification. The principal investigator and two other members of the research team developed the scoring criteria. The criteria for each discourse were based on key words that were critical towards understanding the context of the discourse. This information was gathered from four individuals who generated 2-3 titles for each complete discourse. For example, several different responses were considered accurate for the discourse entitled "Wedding ceremony": "wedding," "married," "marriage," and "marry." It was expected that participants would vary in the extent of information they provided for each title identification. Because of this, objective criteria consisted of the minimum necessary information required to accurately identify the context of the discourse.

Two independent raters, blind to the purposes and hypotheses of the study, individually scored the title identification responses. For each discourse by Wave and Condition (Titled, Untitled), the raters were provided with (a) each participant's title identification response, (b) the actual correct title for each discourse, and c) the objective scoring criteria for each discourse. The raters were instructed that each, "Response needs at least one of the following [criteria] to be scored as a correct response." A third independent rater subsequently resolved any scoring discrepancies between the two primary independent raters. Like the primary two independent raters, this third independent rater was blind to the study procedures, hypotheses, and identities of the other raters.

The raters scored the title identification responses following each Condition (Titled, Untitled) within each Wave. There were approximately 30 participants per each

Condition in each Wave. Therefore, the two independent raters scored approximately 2,400 title responses per each Condition in each Wave. Raters dummy-coded each title response as correct (“1”) or incorrect (“0”). The independent raters scored the responses within 7-days after receiving them. The two independent raters established a high-level of inter-rater reliability for each Wave (Wave 1-Untitled: r = 0.997; Wave 1-Titled: r = 0.966; Wave 2-Untitled: r = 0.991, Wave 2-Titled: r = 0.964; Wave 3-Untitled: r = 0.993; Wave 3-Titled: r = 0.949; Wave 4-Untitled: r = 0.966; Wave 4-Titled: r = 0.950).

After completing data collection for Study 1, the research team reviewed participants’ qualitative responses and determined that certain responses were “acceptable” but were not included in the original scoring criteria. For example, the discourse “Having a Picnic” could also reasonably be about a camping trip. Therefore, the research team modified the objective criteria for seven of the 80 discourse and asked the raters to re-score all items originally scored as “incorrect” for the seven discourses from all Waves. The raters re-scored these items only using the new criteria for these seven discourses. The third independent rater again solved any scoring discrepancies between the two primary independent raters.

2.4. Results and stimuli selection for Study 2

Discourses were retained for Study 2 if less than or equal to 14% of participants in the Untitled condition identified the titles of the discourses after reading the first complete two sentences (Waves 1-3). In addition, discourses were retained in which at least 72% of participants in the Untitled group identified the title after reading the full discourse (Wave 4). Also, cloze probability of words in the Untitled condition were maintained below 20%, and cloze probability for Critical Word 3 in the Titled condition was greater than or equal to 45%. Twenty-five of the 80 discourses met inclusion criteria based on title identification accuracy and cloze probability results. Appendix A reports the descriptive statistics for the 25 discourses, including cloze probability for each of the critical words. Appendix B reports the title identification accuracy for the 25 discourses. The data reported below comes from these 25 discourses.

2.4.1. Cloze probability results

A 2 (Condition: Titled, Untitled) x 3 (Critical Word: 1, 2, 3) between-subjects ANOVA examined the extent to which the cloze probability of critical words varied by Condition. Cloze probability results appear in Table 2.2. A main effect of Condition showed that, across critical words, cloze probability was higher in the Titled condition (30.38 ± 31.94) compared to the Untitled condition (3.00 ± 4.67), $F(1, 150) = 170.11$, $p < .001$. There was also a main effect of Critical Word, $F(2, 150) = 91.12$, $p < .001$. Main effects were superseded by an interaction between Critical Word and Condition, F(2,

Table 2.2

Differences in Cloze Probability (%) Between Conditions for each Wave Averaged Across the 25 Discourses

	Condition (M ± SD)			
Critical Word	Untitled	Titled	F(1, 48)	p-value
1	1.10 ± 2.92	11.57 ± 18.84	7.55	.005**
2	2.45 ± 5.30	11.57 ± 18.02	5.90	.019*
3	5.46 ± 4.50	67.99 ± 15.96	355.33	< .001***

Note. ***p < .001, ** p < .01, *p < .05

150) = 70.16, $p < .001$.

Pairwise comparisons indicated that cloze probability was larger in the Titled group compared to the Untitled group for Critical Word 1, $F(1, 48) = 7.55$, $p = .005$, Critical Word 2, $F(1, 48) = 5.90$, $p = .019$, and Critical Word 3, $F(1, 48) = 355.33$, $p < .001$. Therefore, in the Titled condition, although mean cloze probabilities were low to Critical Word 1 (11.57 ± 18.84) and Critical Word 2 (11.57 ± 18.02), clozure was significantly higher than the Untitled groups for these words. Cloze probabilities in the Untitled condition did not significantly vary by Critical Word, $F(2, 144) = 0.75$, $p = .473$. Cloze probabilities between Critical Words in the Titled groups significantly varied, $F(2, 144) = 160.53$, $p < .001$. As shown in Figure 2.2, cloze probabilities in the Titled group were significantly larger to Critical Word 3 (67.99 ± 15.96) compared to Critical

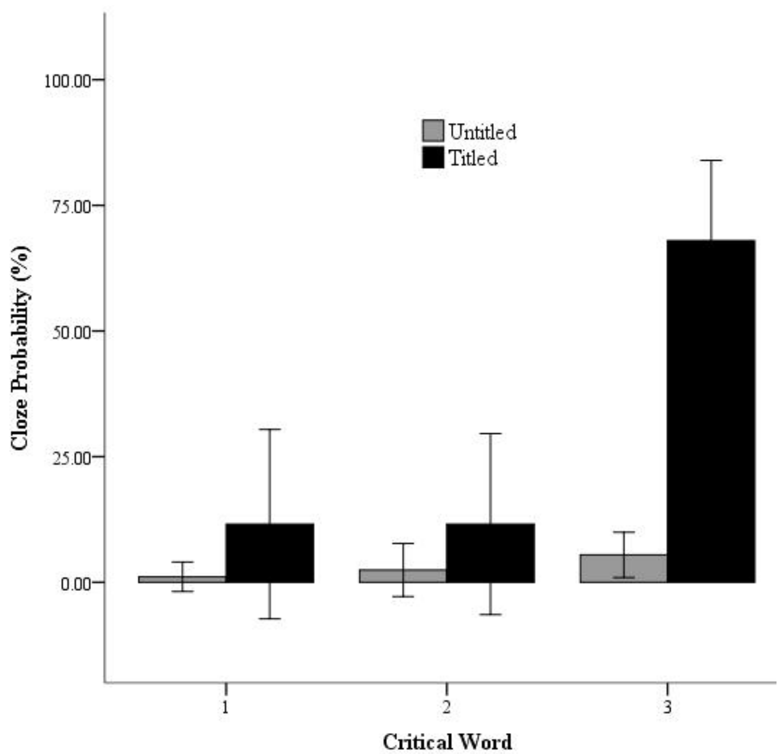

Figure 2.2. Differences in Cloze Probability (%) between Untitled and Titled Conditions for the Three Critical Words

Word 1 (11.57 ± 18.84), $F(1, 48) = 130.50$, $p < .001$ and Critical Word 2 (11.57 ± 18.02), $F(1, 48) = 137.28$, $p < .001$. There was no difference between Critical Words 1 and 2 for the Titled group, $F(1, 48) = 0.00$, $p = 1.0$. In summary, cloze probability was higher in the Titled condition compared to the Untitled condition across all three critical words, but this difference was the largest to Critical Word 3. Cloze probability was equal between critical words in the Untitled group. Therefore, a significantly higher cloze probability to Critical Word 3 in the Titled group compared to Critical Words 1 and 2 was established. This indicates that Critical Word 3 was more semantically related to the contexts of the discourses than Critical Words 1 and 2.

2.4.2. Title identification results

Because participants in the Titled group were provided with the titles of the discourses, the focus of this analysis was on participants' title identification in the Untitled group. Title identification results appear in Table 2.3. A One-Way ANOVA demonstrated a significant effect of Wave on title identification, $F(3, 100) = 2662.74$, $p < .001$. As shown in Figure 2.3, title identification accuracy was significantly larger for Wave 4 (82.22 ± 5.86) compared to Wave 3 (6.80 ± 4.56), $F(1, 48) = 2579.38$, $p < .001$. Title identification accuracy was also significantly greater for Wave 3 than Wave 2 (0.52 ± 1.53), $F(1, 48) = 42.67$, $p < .001$. Therefore, title identification accuracy was significantly higher in the Untitled group after reading the full three-sentences of each discourse compared to the first two sentences. Accuracy was also higher after reading the first two sentences than the first sentence alone. Appendix B demonstrates the descriptive data for title identifications for the 25 discourses.

Table 2.3

Differences in Title Identification Accuracy (%) for Untitled Waves Averaged Across the 25 Discourses

Pairwise comparison (M ± SD)		F(1, 48)	p
Wave 1	**Wave 2**		
0.82 ± 1.48	0.52 ± 1.53	0.50	.484
Wave 2	**Wave 3**		
0.52 ± 1.53	6.80 ± 4.56	42.67	< .001***
Wave 3	**Wave 4**		
6.80 ± 4.56	82.22 ± 5.86	2579.38	< .001***
Wave 4	**Wave 2**		
82.22 ± 5.86	76.98 ± 10.65	4.74	< .001***
Wave 1	**Wave 3**		
0.82 ± 1.48	6.80 ± 4.56	38.93	< .001***
Wave 1	**Wave 4**		
0.82 ± 1.48	82.22 ± 5.86	4530.95	< .001***

Note. ***p < .001, ** p < .01, *p < .05

In summary, a 2 (Condition: Titled, Untitled) x 4 (Wave: 1, 2, 3, 4) between-groups sequential data collection study was conducted to develop a set of three-sentence narrative discourses. The themes of these discourses were ambiguous without a descriptive title. Participants were placed into one of eight groups. Participants read 80 discourses that ranged between 1 and 3 sentences depending on their group. Half of the groups read the discourses with a title. First, participants provided a descriptive title for each discourse based on the sentences read. Second, participants read the same

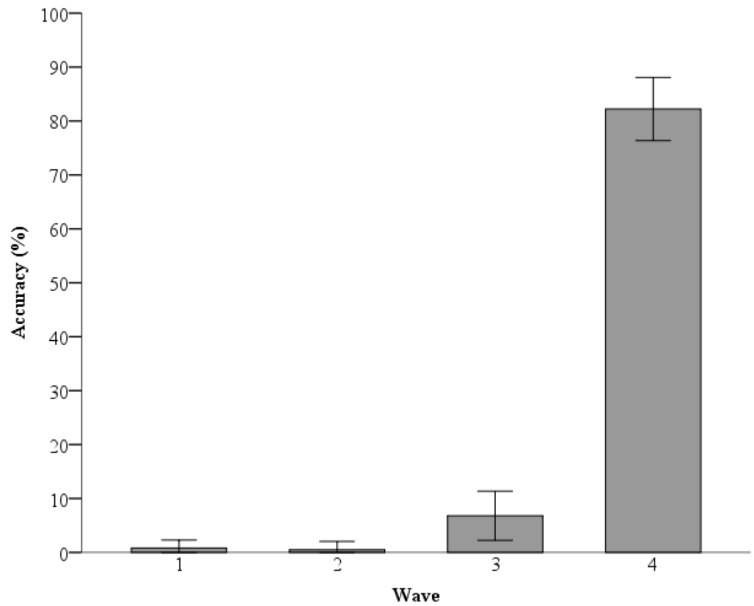

Figure 2.3. Differences in Title Identification Accuracy (%) between Untitled Waves

sentence(s) from the current discourse and then the next sentence occurring in the discourse. The last word of this sentence was missing. Participants completed the sentence with the word they felt fit best. The research team developed the discourses in sequential order, such that the cloze probability from sentence 1 (Wave 1) was first established. Participants in Wave 2 read two-sentence discourses; they read the first sentence (established in Wave 1) and then made cloze judgments on the final word of the second sentence. The same procedure was performed for Wave 3.

From the original set of 80 discourses, 25 were retained that met the following criteria: (a) low-cloze probability (< 20%) for Critical Words 1-3 in Untitled condition; (b) high-cloze probability for Critical Word 3 in Titled condition (> 45%); (c) low title identification (< 14%) in the Untitled group after reading first two sentences (i.e., Waves 1-3); (d) high-title identification (> 72%) in the Untitled group after reading three full sentences (i.e., Wave 4). Data from these 25 discourses indicated that, in the Titled groups, cloze probability was significantly greater to Critical Word 3 compared to Critical Words 1 and 2. This indicates that Critical Word 3 was more semantically related to the contexts of the discourses than Critical Words 1 and 2. In addition, participants in the Untitled groups' identified the titles significantly more often after reading the full discourses compared to only reading the first two sentences. Title identification accuracy was also greater after reading the first two sentences than the first sentence alone.

Therefore, the presentation of Critical Word 3 (high cloze probability in Titled group, not Untitled group) led participants to identify the topics of the 25 discourses when they were not accompanied by a title. This is because participants in the Untitled condition in Wave 4 (read full discourses) were significantly more accurate in identifying the contexts than Untitled participants in Wave 3 (read first two sentences). Unexpectedly, participants in the Untitled groups were also more accurate in identifying the contexts of the discourses after reading sentences 1 and 2 than reading the first sentence alone.

CHAPTER 3

STUDY 2: CHANGES IN THE N400 AND P600 EVENT-RELATED BRAIN POTENTIALS DURING COMPREHENSION OF AMBIGUOUS NARRATIVE DISCOURSES

This chapter reviews the methods used to examine changes in the N400 and P600 ERPs during the comprehension of the 25 contextually ambiguous discourses from Study 1. Researchers suggest that the N400 represents an automatic process of lexical-semantic retrieval that relates to the expectancy of a word within a context (Kutas & Federmeier, 2000). However, the P600 may represent a post-lexical process of context updating (Brouwer et al., 2012). If so, N400 amplitudes to the 25 Untitled discourses in the current study should not fluctuate to Critical Word 3. This is because we established a low cloze probability of Critical Word 3 in the Untitled condition. On the other hand, the P600 should be increasingly larger from Critical Word 1 to Critical Word 2 and Critical Word 3. This is because we established that participants' title identifications in the Untitled condition were significantly more accurate after reading Critical Word 2 than Critical Word 1 and after reading Critical Word 3 than Critical Word 2. Taken together, these ERP findings would indicate that the N400 is sensitive to automatic lexical processes that may be facilitated by a context (i.e., lexical-semantic retrieval; Kutas & Federmeier, 2000). However, the P600 is sensitive to one's identification of changes of a context—i.e. discourse context updating.

3.1. Participants

A sample size of 28 subjects is sufficient to test the 3 (Critical Word: 1, 2, 3) x 2 (Group: Titled, Untitled) mixed-subjects interaction on the amplitude of the N400 and P600 ERPs. This estimate is based on a power analysis with a moderate effect size ($f = 0.25$), alpha of 0.05, 0.80 power, a 0.50 correlation among repeated measures, and a nonsphericity correction value of 1.0 (G*Power 3.1). We oversampled this estimate and tested 43 undergraduate students. Four participants were removed from data analysis because of excessive artifacts during the ERP task (e.g., eye blinks, eye movements). Data was incomplete for an additional two participants because of equipment malfunctions during EEG acquisition. As a result, the final sample consisted of 37 participants. All participants were enrolled in at least one psychology course at the University of Nebraska-Lincoln. Sample characteristics and demographics are reported in Table 3.1. The ages of participants ranged from 18.33 to 22.91 ($M = 20.07$, $SD = 1.22$). The Titled and Untitled groups did not significantly differ in age, $F(1, 35) = 0.95$, $p = .337$, or years of education completed, $F(1, 35) = 0.92$, $p = .344$. Years of education completed ranged from 13 to 16 ($M = 14.0$, $SD = 1.03$). Males and females were equally represented in both groups.

3.2. Recruitment

Participants were recruited through the UNL Psychology Department's Mass Screening study and SONA subject pool. Undergraduate students enrolled in Psychology classes were eligible to complete an online survey for SONA credit that they could apply

Table 3.1

Study 2 Descriptive Statistics of Sample by Group

	Titled (n = 18)	Untitled (n = 19)
Age (M ± SD)	20.27 ± 1.24	19.88 ± 1.21
Gender[1]	9, 9	10, 9
Education (M ± SD)[2]	14.17 ± 1.10	13.84 ± 0.96

Note. [1] Females, Males; [2] Years of formal education completed

to course requirements or extra credit as designated by their instructor. A subset of participants who completed the Mass Screening study also provided contact information to be consented for future research opportunities. The research team recruited those participants who met the primary inclusion criteria (as discussed below). Researchers called and/or texted potential participants. During this initial contact, potential participants were briefly informed of the study procedures and time commitment involved in participating in the study. Interested participants could also contact the research team directly to sign up to participate. Participants enrolled in Psychology courses at UNL received 5 SONA research credits for their participation (1 credit per 30 minutes of participation).

3.3. Eligibility and exclusion criteria

Initial exclusion was determined based on participants' responses on the Mass Screening Form or immediately prior to data collection procedures. Participants were excluded from participating if they endorsed any of the following conditions: cochlear implants, language and/or speech disorder, reading disorder, learning disorder, epilepsy, brain tumor, brain surgery, central nervous system disease, shrapnel or neurostimulator in body, sleep disorder, diagnosed psychiatric disorder, attention-deficit hyperactivity disorder, or developmental disorder (e.g., autism spectrum disorder). Subjects were also excluded from participating if they experienced a blow to the head that caused loss of consciousness and/or posttraumatic amnesia, or a diagnosed concussion/brain injury within the last 7 years. Subjects who did not endorse any of these criteria were invited to participate. Participants were further excluded if they consumed alcohol/illicit drugs in the past 24 hours and/or received less than 4 hours of sleep the night before. A vision screening was administered to establish normal/corrected-normal visual acuity of 20/30 or better.

The vision screening included the standard Snellen test of visual acuity. The left and right eyes were tested separately. Participants wore corrective lenses/contacts if necessary. Administration of this screener followed standard procedures. In brief, participants stood 20 feet (i.e., 6 meters) from the 22" x 11" Snellen chart. Participants read the optotypes (English letters) on each individual line, starting from the top and going to the bottom. The first optotype was 88.6 millimeters, with the letters getting progressively smaller after each line. The experimenter stopped the vision screener when the participant incorrectly read three or more optotypes on a single line. Visual acuity was recorded as that at the line prior to the line in which the three mistakes were made.

3.4. Study procedures

Following prescreening, participants completed the Ambiguous Narrative Discourse ERP task, the Reading Span Task (Daneman & Carpenter, 1980; Kane et al.,

2004) and the Boston Naming Test Second Edition (Kaplan, Goodglass, & Weintraub, 2001). The ERP task was administered first, followed by the behavioral tests. The order of the behavioral tests was counterbalanced within each Group (Titled Discourse, Untitled Discourse).

3.4.1. The Reading Span Task

Participants completed an automated version of the Reading Span Task (RST; Daneman & Carpenter, 1980; Kane et al., 2004) to assess working memory capacity during reading. Similar to other research (e.g., Kolk et al., 2003; St. George et al., 1997; Vos, Gunter, Schriefers, & Friederici, 2001), the purpose for administering the RST was to see how working memory abilities moderated ERPs elicited during discourse comprehension (i.e., N400, P600). This task provides a measure working memory capacity while reading but is not a method of assessing reading capabilities per se. Importantly, the RST taxes the storage and processing aspects of working memory as compared to other working memory tasks such as the digit span that only tax the storage process. Therefore, the RST provides a realistic measure of the ongoing storage and processing aspects of working memory that individuals use during reading.

Daneman and Carpenter (1980) developed the Reading Span Task to assess working memory capacity. In contrast to traditional measures of working memory (e.g., digit span) reading span is highly related to reading comprehension and Verbal SAT scores (Daneman & Carpenter, 1980), suggesting that the RST is a criterion-related measure of working memory capacity while reading. Kane and colleagues (2004) modified the original Reading Span Task. In contrast to the original version in which participants recalled the last word in each sentence, the researchers suggested that the modified version eliminated individual differences in reading abilities.

Participants read aloud groups of sentences (i.e., “a set”) that were visually presented on a computer monitor. Sentences were comprised of American English words and were syntactically correct. Half of the sentences contained semantic anomalies such that these sentences were nonsensical. Each sentence was between 10-15 words in length. The number of sentences per set ranged between three and seven. Between each sentence within a set, participants were visually presented with one of 12 individual letters for 1000 ms. There were three sets for each of the five Levels (3-7) for a total of 75 sentences and 75 letters. Participants were instructed to read the sentence as quickly as possible and indicate if the sentence made sense by clicking “True” or “False” on the computer screen. Immediately after, participants were presented with the letter on the computer screen, followed by the next sentence in the set. After each set, participants saw a 4x3 letter matrix and indicated the order in which the individual letters were presented. The final outcome measure was the number of recalled items across the 15 sets and ranged from 0-75 (Redick et al., 2012).

This test was automated and guided by the participant (Unsworth, Redick, Heitz, Broadway, & Engle, 2009). The automated version of the RST demonstrated a large test-retest reliability of 0.82 when the partial scoring method was used, such as in the present study. That is, participants received partial credit for correctly recalled letters within a set, even if they did not correctly recall all letters within a set. The partial scoring method also demonstrated high intraclass correlation coefficients between 0.86-0.88 (Redick et al., 2012).

There were three practice stages prior to the experimental trials. In the first stage, participants completed the letter span portion of the task: They were presented with two letters, one after another, and then recalled the order in which they were presented given the 4x3 matrix. In the second stage, the participants completed the sentence-processing portion only: They read three sentences and indicated if each sentence was "True" or "False." Participants were instructed to read these sentences as fast as possible. The stimulus presentation program (E-Prime 2.0) calculated the average time it took the participants to read the sentence and make their response. The average value ±2.5 SDs was used in the experimental trials to determine the time window during which participants should realistically read the sentence and make a response. If the participant did not make a response within this time window, the experiment proceeded to the next screen and the trial was recorded as an error. The third stage mimicked the experimental trials. Participants completed three trials of each practice stage prior to beginning the experimental trials.

3.4.2. The Boston Naming Test—Second Edition

Confrontation naming concerns the ability to use language to identify a stimulus and requires intact lexical-semantic retrieval, the ability to retrieve linguistic information of a stimulus from semantic memory. The Boston Naming Test—Second Edition (Kaplan, Goodglass, & Weintraub, 2001) was used because it is a clinically validated and normed assessment of confrontation naming abilities.

The Boston Naming Test is a subtest included in the Boston Diagnostic Aphasia Examination, is normed for adults aged 18-79 years of age, and takes approximately 5-10 minutes to administer. To administer the Boston Naming Test, a trained researcher presented up to 60 individual black and white line drawings to the participant. The participant verbally identified the name of each individual object as quickly as possible. The set of line drawings increased in level of difficulty. For example, the third image is of a pencil, whereas the fiftieth image is a protractor.

In accordance with the recommended administration procedures for individuals over the age of 10 and adults without aphasia (Kaplan, Goodglass, & Weintraub, 2001), the researcher began the test at picture 30 (harmonica). To begin, the experimenter provided the participants with the following instructions: "I am going to show you a series of pictures. Please tell me the name of each of these pictures as quickly as you can. If you are not sure, just give your best guess." Participants' responses were audio recorded using a digital audio recorder.

Starting from item 30, the researcher worked upwards, showing the participant each subsequent line drawing one at a time. The administrator showed the remaining individual items sequentially unless the participant made an error prior to picture 38 (harp). In this event, the administrator returned to item 29 and worked backwards until the participant made eight consecutive correct responses (e.g., items 29-22). When working backwards, correct responses were provided for all items that preceded the starting point in which the participant made eight consecutive correct responses (e.g., item 29). If the participant made eight consecutive responses when working backwards, the administrator then showed the next item after the participant's initial incorrect response. Following, the administrator showed the remaining items in sequential order. For example, if the participant made her first mistake at image 36 (cactus), the administrator would skip downwards to item 29. If the participant correctly identified eight consecutive items (i.e., items 29-22), the administrator would next show item 37. If the participant made additional errors on items 54 (tongs) and 60 (abacus), her final score would be 57. The task finished when the participants made her response for the last item (abacus). Participants were given up to 20 seconds to make a response or less if the participant indicated he/she did not know the word. The maximum score was 60.

The administrator did not provide any stimulus cues. The same researcher administered the Boston Naming Test to all participants. Following completion of the assessment, a trained researcher scored each individual item as correct or incorrect. Outcome measures include the number of correctly identified items and the average latency to name each item. Using Audacity 2.1.2 Software, the latency of each participant's responses (0.1-20s) was determined as the offset of the administrator's dictation of the item number and the onset of the participant's dictation of the name of the image. Latencies were averaged for all correct responses. Latency was not recorded for incorrect responses.

3.4.3. Event-related potential recording: Ambiguous Narrative Discourse Task

3.4.3.1. Stimulus set

Stimuli consisted of 25 three-sentence paragraphs (Study 1). As reviewed in Chapter 2, the cloze probability of Critical Word 3 was significantly higher than Critical Words 1 and 2 in the Titled condition. However, in the Untitled condition, cloze probabilities did not differ between Critical Words. These findings established that Critical Word 3 was highly semantically related to the contexts of discourses. In addition, participants' identification of the contexts of discourses increased after reading Critical Word 2 (6.80% ± 4.56) and again after reading Critical Word 3 (82.22% ± 5.86). This established that greater discourse context updating occurred after reading Critical Word 2 and Critical Word 3. Researchers recommend that 25-40 trials for each condition are sufficient to maximize the signal-to-noise ratio of averaged ERPs (Kutas, DeLong, & Kiang, 2011; Kutas & Van Petten, 1994; Van Berkum, 2004). Because participants would only view discourses in one condition (Titled Discourse, Untitled Discourse), a stimulus set of 25 discourses was sufficient to produce averaged ERPs with sufficient signal-noise ratios.

A number of parameters are thought to influence the timing and morphology of the N400 and P600 ERPs—most of which are not relevant to the purposes of this study. Consequently, most these parameters, if not controlled, may pose confounds to the present study. In the following sections, the methods employed to control these parameters are discussed.

3.4.3.1.1. Sentence position, word class, and number of words

All Critical Words were open-class nouns and occurred in the sentence-final position. The position of a word within a sentence may influence the amplitude of the N400 (e.g., Kutas et al., Besson, 1988; Van Petten, 1993; Van Petten & Kutas, 1990). Research suggests that open- and closed-class words generate qualitatively different ERPs. Specifically, open-class words generate a typical N400 effect that is maximal at centroparietal electrode channels, whereas open-class words generate an earlier peak around 280 ms that is maximal at anterior electrode channels (Kutas et al., 1988; Neville, Mills, & Lawson, 1992). As shown in Table 3.2, averaged across paragraphs, the three sentences did not differ in number of words, $F(2, 74) = 0.50$, $p = .661$). Appendix A provides the descriptive statistics for each of the individual 25 discourses.

Table 3.2

Differences in Sentence- and Critical Word Parameters Across the 25 Discourses

	Sentence (M±SD)			Min-Max	F	p
	1	2	3			
Words	7.16 ± 1.40	7.20 ± 0.96	6.92 ± 0.76	5-10	0.50	.661
CW Length (letters)	5.76 ± 2.20	6.56 ± 2.29	5.44 ± 1.64	3-11	1.95	.149
CW Syllables	1.52 ± 0.82	1.80 ± 0.87	1.48 ± 0.65	1-4	1.23	.298
CW Concreteness	4.38 ± 0.52	4.48 ± 0.46	4.70 ± 0.31	3.17-5	3.33	.041*
CW Frequency	2559.68 ± 4040.94	3367.08 ± 6816.19	2362.88 ± 4844.27	2-26214	0.25	.782

Note. CW = Critical Word. *p < .05.

3.4.3.1.2. Word frequency

As shown in Table 3.2, Critical Words did not vary in word frequency. The frequency with which words occur in a language influences word recognition (for review see Cortese & Balota, 2012; Forster & Chambers, 1973; Hauk & Pulvermüller, 2004) and N400 amplitudes (Rugg, 1990; Van Petten & Kutas, 1990). Word frequency was determined using the SUBTLEX U.S. Corpus (Brysbaert & New, 2009), which contains 51.0 million words from 8,388 U.S. films and television shows from the years 1900-2007. The size of this corpus is substantially larger than other American English corpora, such as the Brown University Corpus (1.0 million words; Francis & Kučera, 1982; Kučera & Francis, 1967). Brysbaert and New (2009) argued that language use in television and film is more realistic to everyday communication and does not suffer from the same artificial lack of variation as word use in written texts (e.g., Francis & Kučera, 1982; Kučera & Francis, 1967).

3.4.3.1.3. Word repetitions

Research also suggests that both word recognition and the amplitude of the N400 are modulated by repetitions of previously presented words. Specifically, N400 amplitudes are smaller to repeated words or non-words compared to unique words or non-words (Deacon et al., 2004; Rugg, 1985). Therefore, Critical Words in the current study were minimized for repetition effects. Only one Critical Word was repeated. This word ("car") was used in the sentence-final position for sentence 3 in two different paragraphs. Because all discourses were randomized, we presumed that there would be no repetition effects (i.e., smaller N400 amplitudes) to the second presentation of this word. To further avoid repetition effects, titles of the discourses did not contain any of the Critical Words.

3.4.3.1.4. Semantic priming, leading sentences, and clichés

Semantic associations between word pairs influences word recognition (Evett & Humphreys, 1981) and N400 amplitudes (Deacon et al., 2000; Rolke et al., 2001). The degree to which a word is semantically related within a local sentence also influences the N400 (Kutas & Hillyard, 1980), and there may be an additive effect of sentential and lexical priming effects on the N400 (Van Petten, 1993). Therefore, the stimulus set was carefully constructed to avoid potential local sentential priming effects that would facilitate lexical-semantic retrieval of Critical Words above and beyond the contextual support provided by the Title (in the Titled Discourse group). Therefore, this stimulus set was designed to limit priming effects due to local sentential context but, rather, manipulate contextual priming due to the presentation of a title.

To avoid semantic priming effects from non Critical Words, sentences were also constructed to avoid clichés (Bloom & Fischler, 1980), such as "The apple does not fall far from the tree." For the same reason, strings of words that commonly occur together to make a well-known phrase were avoided (e.g., "peanut butter and jelly"; Forster, 1981). Importantly, the N400 is modulated by the expectancy of a word within a given context. As just reviewed, it has been suggested that a prior context serves as a prime for the lexical-semantic retrieval (indexed by the N400) of subsequent words (Deacon et al., 2000; Kutas & Hillyard, 1980; Rolke et al., 2001). Therefore, by presenting subjects with clichés or leading phrases, participants may be able to anticipate the Critical Word several words prior to its actual presentation. Paragraphs were carefully controlled so that they did not contain any semantic priming in the local sentence, leading sentences, or clichés.

3.4.3.1.5. Syntax

Researchers disagree on the functional significance of the P600 ERP as a marker of syntactic repair (Friederici et al., 2004; Hagoort et al., 1993; Kotz et al., 2003; Osterhout & Holcomb, 1992) or, for example, discourse context updating (Brouwer et al., 2012). The hypothesis in the present study is that the P600 is a marker of discourse context updating. Therefore, the stimulus set was constructed such that sentences ranged in syntactic complexity. The majority of sentences were constructed in the subject, verb, object word order (98.17%). There were no intended operations on syntax (Bloom & Fischler, 1980).

3.4.3.1.6. Concreteness

Concrete words (e.g., "door") may generate more negative N400 amplitudes compared to abstract words (e.g., "truth") at frontal electrode channels (e.g., Barber, Otten, Kousta, & Vigliocco, 2013; Holcomb, Kounios, Anderson, & West, 1999; Kounios & Holcomb, 1994; Xiao et al., 2012). However, this concreteness effect may be specific to anomalous words in a sentence, not coherent words (Holcomb et al., 1999). Critical Words' level of concreteness in the current study was established using norms for 63,039 English word lemmas (Brysbaert, Warriner, & Kuperman, 2014). All words in this study were at least "moderately concrete," as determined by average independent ratings that were greater than 3 on a 5-point Likert scale from 1 ("Abstract") to 5 ("Concrete"; Brysbaert et al., 2014). The concreteness of Critical Words ranged from 3.17 to 5.0 (4.52 ± 0.45). As shown in Table 3.2, concreteness varied by Critical Word, $F(2, 74) = 3.33$, $p = .041$. Follow-up comparisons indicated that concreteness was higher to Critical Word 3 (4.70 ± 0.31) compared to Critical Word 1 (4.38 ± 0.52), $F(1, 48) = 6.70$, $p = .013$.

3.4.3.1.7. Word length

There is a lack of convergence on whether word length influences language processing. This may be due to differences in the nature of the stimuli (e.g., lexical decision vs. normal prose reading), ERP analysis procedures (e.g., peaks at isolated electrode channels: Pratarelli; 1995, Van Petten & Kutas, 1990; voltage across all channels: Hauk & Pulvermüller, 2004), or other unexpected lexical factors specific to each study. However, taken together, it is unlikely that word length has a null influence on language comprehension. Even if word length effects are not exhibited through the morphology of the N400 component, they may influence earlier sub-lexical processing (e.g., Hauk & Pulvermüller, 2004). Although such early-evoked potentials (e.g., P100, N100) are not the primary focus in the current study, they may be processed and analyzed for future research. Given the possible interpretations of word-length effects previously reviewed and the recommendations of such work (Hauk & Pulvermüller, 2004; Pratarelli, 1995), word length between Critical Words was controlled in the current study. Length of Critical Words ranged from 3-11 letters (5.92 ± 2.09) and did not vary between Critical Words, $F(2, 74) = 1.95$, $p = .149$ (see Table 3.2).

3.4.3.1.8. Number of syllables

The focus of the present study was to mimic, as best as possible, commonly encountered words while reading. For this reason, Critical Words were mono- or polysyllabic and ranged from 1 to 4 syllables (1.60 ± 0.79). The number of syllables for each word was determined from the English Lexicon Project (Balota et al., 2007), an open-source database of over 40,000 words that were normed during speeded naming and lexical decision tasks. As shown in Table 3.2, the number of syllables did not vary between Critical Words, $F(2, 75) = 1.23$ $p = .298$.

3.4.3.1.9. Semantically interesting

The contexts of the discourses in the current study concerned everyday life activities common to college students. Example topics included a wedding ceremony, a graduation, taking a driving test, and going to the movies. The comprehension of these discourses relied on a certain level of background information regarding each theme. For example, to understand the theme of a wedding, participants were expected to know that wedding ceremonies often include flowers and a bride. It was assumed that all participants were familiar with contexts because of personal experience and/or media exposure of events.

The discourse contexts varied to maintain participants' interest throughout the study and maintain their engagement (Nieuwland & Van Berkum, 2006). This was particularly important for several reasons. First, focusing participants' attention was important considering the length of the study (approximately 60 minutes). Therefore, the contexts were designed to be semantically interesting to minimize fatigue effects. Second, all sentences were coherent. This lies in contrast to typical N400 paradigms that include a semantically anomalous word (e.g., Kutas & Hillyard, 1980, 1983).

Each discourse was constructed such that the subject was non-specific and carried no contextual information. For example, the subject of the "Wedding Ceremony" was "the man." Other examples of subjects included, "the woman," "the girl," "the boy," and "the family." The subject was always clear such that the presentation of a title facilitated the understanding of the context, not the subject. All contexts were also constructed to be gender neutral. The gender neutrality of the discourses ensured that male and female participants were equally familiar with the contexts. For example, "the girl" in the first sentence of the context "At the Zoo," could be simply replaced with a male subject. Because going to the zoo is a gender-neutral event, the gender of the subject did not create an unexpected coherence anomaly. The subjects of the sentences were also kept general to avoid names that could potentially be gender biased. The use of vague subject identifiers also controlled for any lexical-semantic associations that a participant may have formed based on prior exposure to certain names. These extraneous lexical-semantic associations may influence attentional bias across paragraphs.

As discussed in prior sections, the stimulus set in the current study was constructed to be coherent and ecologically valid. Such a stimulus set is invariably critical towards beginning to understand how the brain supports the complex integration of bottom-up and top-down processes that are necessary for understanding naturalistic connected speech (Kintsch, 2005).

It was particularly important to keep participants engaged because they did not make a behavioral response. As reviewed in Van Berkum (2004), a behavioral response may place unanticipated attentional demands on a task that are unrelated to the research question at hand. Regan (1972, p. 139) also commented that ERPs recorded to a behavioral response may, in fact, be due to the motor response itself, not the psychological construct of interest (e.g., attention). Said otherwise, the purpose of the task was to create a naturalistic environment that simulated language comprehension. To do so, participants were simply asked to read the discourses for comprehension; there was no behavioral response during EEG recording (e.g., Burkhardt, 2006, 2007; Federmeier & Kutas, 1999; Nieuwland & Van Berkum, 2005; St. George et al., 1994, 1997; Van Berkum et al., 1999; Yurchenko et al., 2013). To encourage participants to maintain their attention throughout the task, participants were asked to identify the title immediately after reading each discourse.

The task remained semantically interesting and maintained participants' attention through the inclusion of 28 "filler" discourses. These filler discourses were intended to distract participants from identifying, and thus automatically expecting, the critical manipulations of the study (see Van Berkum, 2004 for review). There were four types of filler discourses (seven per type). These 28 filler discourses were chosen from the 55 paragraphs from Study 1 that did not meet inclusion criteria (i.e., cloze probability, title identification) for the current study. These discourses were modified from their original form to create the filler versions.

For Type 1 Fillers, sentence 3 and sentence 1 were exchanged, such that the higher cloze-probability word (Study 1) concluded the first sentence. Sentences 2 and 3 were flipped for Type 2 Fillers. Type 1 and Type 2 Fillers were employed to draw participant's attention from Critical Word 3 as the primary locus of manipulation. For Type 3 Fillers, all sentences remained in their original order, but Critical Word 3 was replaced with a semantic anomaly (Kutas & Hillyard, 1980). All semantic anomalies were open-class nouns and were grammatically correct sentence endings. However, the anomaly did not fit with the theme of the discourse (See Appendix C for examples). For Type 4 Fillers, Critical Word 2 appeared in all capital letters (Kutas & Hillyard, 1980). This physical anomaly was intended to draw participant's attention from the semantic aspects of the study. Similar to the experimental discourses, participants read the filler discourses with or without the title depending on their Group (Titled Discourse, Untitled Discourse).

3.4.3.2. Stimulus presentation

Participants sat 155 cm from a computer monitor in a dimly lit room. Participants read the 53 discourses (see Appendix C), presented one word at a time. Twenty-eight of these discourses were filler paragraphs. Words at the beginning of each sentence were capitalized. A period occurred at the end of each sentence-final word. No other punctuation occurred within the paragraphs. A benefit of this serial visual presentation (SVP) method is that it avoids overlapping ERPs that are inherent in connected discourse. Participants were instructed to minimize their body motion and eye blinks throughout the experiment. During brief 2- to 3-minute break periods in the study, participants were able to move their arms, legs, and head while remaining seated.

Participants were randomly assigned to the Titled Discourse or Untitled Discourse groups. Participants in the Titled Discourse group, but not Untitled Discourse group, read the descriptive title prior to each discourse. The order in which the paragraphs were presented was random. Prior to beginning the experimental block, participants completed a practice block of four discourses. Two of the practice discourses were coherent. One practice discourse contained a semantic anomaly (Type 3 Filler). One practice discourse contained a physical anomaly (Type 4 Filler). Participants were instructed to read all stories for comprehension. This stimulus presentation sequence is depicted in Figure 3.1.

All words were presented in white Times New Roman font in the center of the screen on a black background. The visual angles of Critical Words in the primary 25 discourses ranged from 0.591 to 1.995 (1.234 ± 0.388). Prior to each paragraph, participants were presented with a screen that read, “Ready?” Next, for those in the Titled Discourse group, the title was visually presented for 3000 ms. After the title

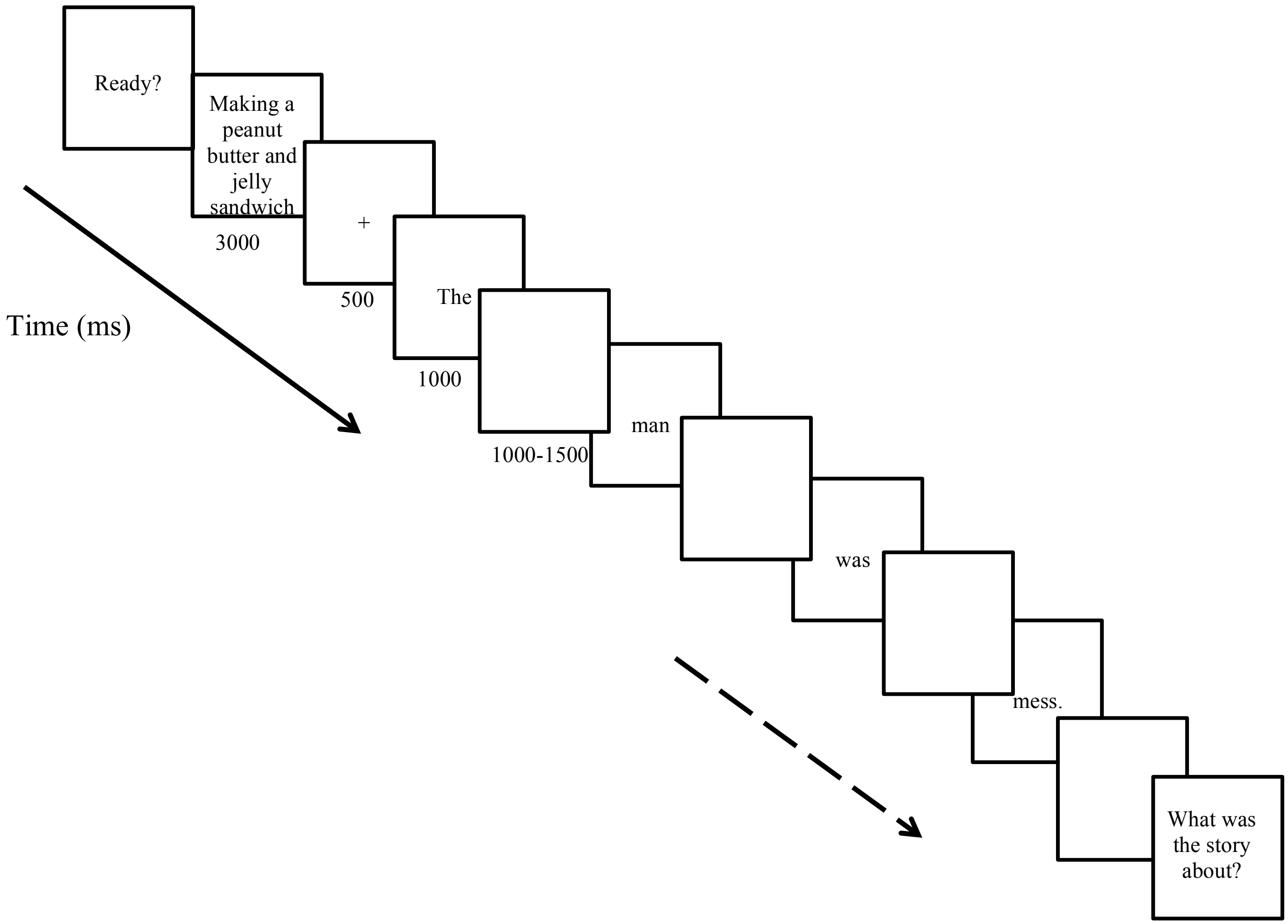

Figure 3.1. Stimulus Presentation Sequence for Ambiguous Narrative Discourse ERP Task

screen, a fixation cross appeared in the center of the computer screen for 500 msec. In the Untitled Discourse condition, this fixation screen occurred after the “Ready?” screen. Each word was then presented for 1000 ms followed by a blank screen (Van Berkum et al., 1999) that was presented between 1000-1500 ms before the presentation of the next word. A period denoted the last word in each sentence. Following the 1000-1500 ms blank screen after the final Critical Word, participants were presented with a screen that said, “End of the story. What was the story about?” Participants verbally responded with their identification of the title of the discourse. A researcher transcribed participants’ verbal responses. This screen was presented until the participant finished his/her verbal response and was ready to continue to the next discourse. Following this screen, the “Ready?” screen appeared to prepare the participant for the next discourse. Testing time for each discourse lasted approximately 1 minute. Participants were given short breaks every 10-15 discourse. During this time, the researchers adjusted electrode channel impedances greater than 60 kΩ.

The time-varying ISI between words (1000-1500 ms) was employed for several reasons. Prior research suggests that amplitude of an endogenous, slow-wave, negative ERP called the contingent negative variation (CNV) reflects the preparation for an upcoming stimulus (Regan, 1972, p. 134). Importantly, this effect is minimized when expected stimuli are presented in random order and at varying time intervals. As reviewed in Regan (1972, p. 141), Näätänen (1967, 1970) presented participants with relevant and irrelevant auditory click stimuli and found that amplitudes were the same in the two conditions when they were presented in random order with varying ISIs. However, when the click stimuli were presented at fixed intervals and in a known order, the ERPs to the relevant stimulus clicks was different when compared to the clicks that were classified as irrelevant (Näätänen, 1967, 1970). Näätänen suggested that the larger amplitude to relevant clicks in the fixed ISI condition may be due to a pre-stimulus preparation effect for the relevant stimulus click; however this contrast in ERP amplitude findings between different methodological designs failed to replicate in the visual modality. This factor is important because if a negative potential precedes a stimulus presentation, it may likely minimize the amplitude of later-occurring positive potential following the actual presentation of the stimulus (reviewed in Regan, 1972, p. 148).

Secondly, the time-varying ISI was designed to reduce the habituation response to similar and successive stimuli (Regan, 1972, p. 134). Importantly, there seems to be a negative relation between the amplitude of ERPs to a stimulus-type and the number of presentations of that stimulus in a testing session. If a person knows when a stimulus will occur (either because the order of conditions is not random or the interval between stimuli is the same), then the ERPs the participant generates may be due to general levels of arousal due to the expected stimulus and not the psychological properties engaged in processing that stimulus (Regan, 1972, p. 139). A time-varying ISI allows for participants to always be alert, rather than only alerting themselves immediately before the stimulus. If the participant cannot predict the time at which a stimulus will occur, then they must maintain a constant state of alertness.

3.4.3.3. Electrophysiological recording and processing

Ongoing electroencephalogram (EEG) was recorded from 256 high-density Ag/AgCl electrodes using Net Station 5.3.0.1 software (Electro Geodesics, Inc., Eugene, OR) and a 1000 Hz sampling rate. Electrode impedances were kept below 60 kΩ at the start of EEG recording. Electrode impedances were adjusted approximately every 10-15 paragraphs throughout testing sessions during participants' breaks. Data were unfiltered during recording. A band-pass filter was applied offline after data collection from 0.1 to 30 Hz (rolloff = 24db/octave). The ongoing EEG was segmented to epochs of 200 ms before and 1000 ms after the onset of the presentation of the Critical Words. Channels were rejected and replaced using spline interpolation from immediately adjacent electrodes. Channels were removed from all epochs if containing a voltage shift greater than 100 μV for greater than 25% of segments. Trials were removed if they contained an artifact (e.g., eye blink, eye movement) during any epoch with an 80 ms moving average. An eye blink was classified as any voltage shift greater than 100 μV. Eye movements were classified as any voltage shift greater than 55 μV. Bad channels and artifacts were also identified and rejected through visual inspection. The same researcher performed all the manual artifact rejection procedures when blinded to experimental conditions as recommended (Keil et al., 2014; Luck, 2014). Correct trials were baseline-corrected with a 200 ms pre-stimulus period and re-referenced to the average reference of all scalp electrode channels. Trials were averaged separately for each participant, Critical Word, and Group (Titled Discourse, Untitled Discourse). The ongoing EEG was downsampled offline to 250 Hz prior to analyses.

3.5. Data analysis

Analyses were performed using SPSS 23.0 (IBM, Chicago, IL, USA). First, a temporal principal components analysis (PCA) identified the factors that characterized the maximum variability in the averaged ERP waveforms. Varimax rotation with a covariance matrix was used to rotate the factors. It was expected that the temporal PCA would characterize factors representing the time windows of the N400 and P600 ERPs. Mean amplitudes within these temporal windows were extracted at electrode clusters of 5-7 electrode channels (median = 6; Figure 3.2). This included clusters around five midline electrode channels (Fz, FCz, Cz, CPz, Pz) and symmetrical left and right lateralized clusters. These electrode clusters correspond to the 10-10 system and are more reliable than recording from single electrode sites (Dien & Santuzzi, 2005). ERP waveforms were averaged for each Critical Word (1, 2, 3) within the Titled Discourse and Untitled Discourse groups. The primary outcomes were the mean amplitude of the N400, P600, ERPs within the time windows derived from the temporal PCA.

To examine Hypothesis 1, N400 mean amplitudes were submitted to a 5 (Cluster: Fz, FCz, Cz, CPz, Pz) x 3 (Laterality: left, midline, right) x 3 (Critical Word: 1, 2, 3) x 2 (Group: Titled Discourse, Untitled Discourse) mixed-subjects General Linear Model (GLM). The same approach was employed to test Hypothesis 2 on P600 mean amplitudes. To examine Hypothesis 3, two separate GLMs were conducted on N400 amplitudes with Boston Naming Test Accuracy and Boston Naming Test Response Time added as continuous moderators. To examine Hypothesis 4, working memory capacity, as determined by the number of correctly recalled words on the Reading Span Task, (Kane et al., 2004) was added as a continuous moderator to the GLM model with P600 mean amplitude as the outcome. Greenhouse Geisser correction was employed to correct for violations of sphericity. Results were significant at $p < .05$.

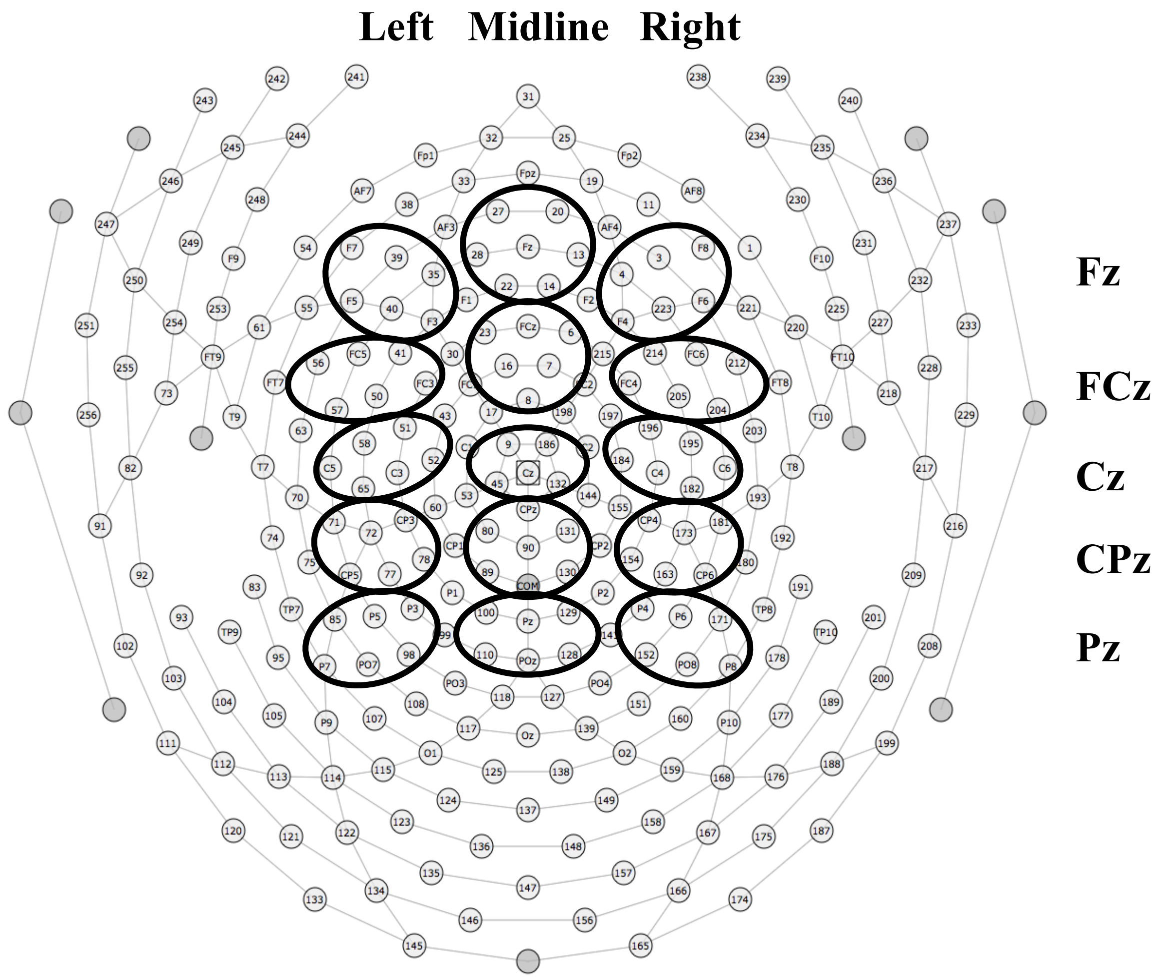

Figure 3.2. Selected Electrode Clusters from Electrical Geodesic, Inc. HydroCel 256 Geodesic Electrode Sensor Net

3.6. Hypotheses

3.6.1. Hypothesis 1

Cloze probabilities of Critical Words were established in Study 1. To summarize, cloze probabilities were higher in the Titled group compared to the Untitled group for all three Critical Words. N400 amplitudes are positively related to the expectancy (i.e., cloze probability) of a word within a sentence (e.g., Kutas & Hillyard, 1980, 1984; Laszlo &

Federmeier, 2009): Amplitudes are more negative to low cloze probability words and less negative to high cloze probability words. Therefore, it is expected that N400 effects in this study would mirror the effects of Condition and Critical Word on cloze probability from Study 1.

Hypothesis 1: N400 amplitudes will not differ between Critical Words for participants in the Untitled Discourse group. For those in the Titled Discourse group, N400 amplitudes will be more negative to Critical Word 1 and Critical Word 2 than Critical Word 3. N400 amplitudes will be less negative for participants in the Titled Discourse group than Untitled Discourse group for all three Critical Words. These effects will be maximal at CPz recording sites.

Support for Hypothesis 1: A significant four-way interaction between Group, Cluster, Laterality, and Critical Word on the N400 amplitude would provide support for Hypothesis 1. It is expected that this effect will be maximal at the right centroparietal electrode cluster (Kutas & Hillyard, 1980; see Kutas & Federmeier, 2000 for review). It is expected that pairwise comparisons for the Titled Discourse group would result in more negative N400 mean amplitudes to Critical Words 1 and 2 than Critical Word 3. No N400 differences between Critical Words are hypothesized for the Untitled Discourse group.

3.6.2. Hypothesis 2

Study 1 established that participants in the Untitled condition were significantly more accurate in identifying the titles of discourses after reading the complete discourses compared to only the first two sentences of the discourses. In addition, title identification accuracy was greater after reading the first two sentences (Wave 3) compared to the first sentence alone. It is presumed that the high cloze probability of Critical Word 3 led to participants' increased ability to identify the titles. Unexpectedly, title identification accuracy increased after reading the first two sentences, but cloze probability to Critical Word 2 did not. Researchers suggest that the P600 ERP reflects the post-lexical process in which the mental model of a story is updated/revised (e.g., Brouwer et al., 2012; Brouwer & Hoeks, 2013; Burkhardt, 2006, 2007). It is expected that our P600 results will map onto our behavioral findings of title identification accuracy established in Study 1. Therefore, it is hypothesized that the mean amplitude of the P600 for those in the Untitled Discourse group will be larger to Critical Word 3 compared to Critical Words 2 and larger to Critical Word 2 than Critical Word 1.

<u>Hypothesis 2</u>: P600 amplitudes will be larger to Critical Words 2 and 3, than to Critical Word 1 for the Untitled Discourse group. P600 amplitudes to Critical Word 3 will also be larger than Critical Word 2 for participants in this group. These ERP findings will correspond to behavioral findings from Study 1, suggesting that participants revise their mental model of the discourses to Critical Word 2 and even more so to Critical Word 3.

<u>Support for Hypothesis 2</u>: A significant four-way interaction between Group, Cluster, Laterality, and Critical Word on the P600 amplitude will provide support for Hypothesis 2. Pairwise comparisons will indicate an effect of Critical Word on the P600 to Critical Words 2 and 3 for the Untitled Discourse group, which is expected to be maximal at midline centroparietal channels (e.g. Brouwer et al., 2012; Brouwer & Hoeks, 2013; Burkhardt, 2006, 2007; Hoeks et al., 2004).

3.6.3. Hypothesis 3

Research indicates that reductions in N400 amplitudes to words that are semantically related to prior words or contexts are due to an automatic process (e.g., spreading activation; Collins & Loftus, 1975). Specifically, researchers suggest that previously presented concepts prime the activation of semantically related concepts in long-term memory (Deacon et al., 2000; Federmeier & Kutas, 1999; Grossi, 2006; Kiefer, 2002; Rolke et al., 2001; Sanford et al., 2011). It has been proposed that the N400 represents the process of retrieving the semantic information of words from long-term memory (lexical-semantic retrieval; Federmeier & Laszlo, 2009; Kutas & Federmeier, 2000, 2011). Confrontation naming tasks, such as the Boston Naming Test, are thought to assess lexical-semantic retrieval abilities.

Hypothesis 3: Confrontation naming performance (accuracy, response time) on the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001) will moderate N400 amplitudes to Critical Words. Specifically, when BNT performance was high N400 amplitudes will significantly vary between Critical Words. However, when BNT performance was low, N400 amplitudes will not vary between Critical Words.

Support for Hypothesis 3: There will be a significant 3-way interaction between Cluster, Group and Confrontation Naming on N400 amplitudes. It is expected that pairwise comparisons will show that N400 amplitudes differ significantly between all three Critical Words when BNT performance is high (+1 SD above the mean). However, N400 amplitudes between Critical Words will not differ when BNT performance is low (-1 SD below the mean). It is expected that these effects will be largest for the Untitled group at centroparietal electrode channels.

3.6.4. Hypothesis 4

During language processing, the ability to integrate bottom-up information within a mental representation of a discourse involves working memory (Clark & Haviland, 1974; Kintsch, 1988; Kintsch et al., 1999; McKoon et al., 1996; Myers & O'Brien, 1998; Ratcliff & McKoon, 1988). Recent research suggests that the amplitude of the P600 is larger to new information presented in a discourse compared to given information (Burkhardt, 2007). Neuroanatomical theories suggests that the P600 ERP is generated by activity in the left inferior frontal gyrus (Brouwer & Hoeks, 2013; Swick, Kutas, & King, 1988; Van Petten & Luka, 2006) which supports semantic working memory processing during discourse comprehension (Cabeza & Nyberg, 2000 for review; Wagner et al., 1997). These findings suggest that increased P600 amplitudes may be related to increased working memory processing to establish/update a mental representation of a discourse (Brouwer & Hoeks, 2013; Brouwer et al., 2012).

Hypothesis 4: Working memory capacity as assessed on the Reading Span Task (Dane & Carpenter, 1980; Kane et al., 2004) will moderate P600 amplitudes to Critical Words. Specifically, when Reading Span is high P600 amplitudes will significantly vary between Critical Words. However when Reading Span is low P600 amplitudes will not vary between Critical Words.

Support for Hypothesis 4: There will be a significant 3-way interaction between Cluster, Group, and working memory capacity on P600 mean amplitudes. It is expected that pairwise comparisons will show that amplitudes will differ significantly between all three Critical Words when working memory capacity is high (+1 SD above the mean). However, P600 amplitudes between Critical Words will not differ when working memory capacity is low (-1 SD below the mean). It is expected that these effects will be largest for the Untitled Discourse group at centroparietal electrode channels.

CHAPTER 4

STUDY 2: RESULTS

4.1. Introduction to results section

Before describing and interpreting results from Study 2, it is worth first reviewing the ways in which the systematic within- and between-group ERP changes in the current study will be interpreted. As Van Berkum (2004) reviewed in-depth, ERPs may be used to examine (a) the sensitivity of a known ERP Component (e.g., N400) to an experimental manipulation (amplitude, measured in microvolts), (b) the relative timing of a language process during comprehension (latency, measures in ms), and (c) the extent to which two or more language processes are identical or non-identical (characterized by differences in polarity, morphology, and/or scalp distributions).

The sensitivity and scalp distribution metrics are most relevant to the current study. For example, a more negative N400 amplitude to Critical Word 1 than Critical Word 2 would indicate that the level of cognitive resources supporting lexical-semantic retrieval (e.g., Kutas & Federmeier, 2000, 2011) were different between the two words. If overlapping scalp distributions (e.g., centroparietal maximum) characterized the N400 amplitudes for Critical Words 1 and 2 (Kutas & Federmeier, 2000, 2011) one would conclude that the neural processes underlying the N400 were similar between words. Instead, however, the N400 effect may have occurred across a more distributed area of electrode clusters to Critical Word 2 than Critical Word 1, but the amplitude, polarity and morphology of the N400s were the same. Therefore, it would be presumed that levels of lexical-semantic retrieval were the same for the two words, but the types of neural resources supporting the processing of each individual word were different. This is because each ERP is characterized by an established polarity (e.g., positive or negative deflection), morphology (shape of the waveform), and scalp distribution (Van Berkum, 2004). Although scalp recordings do not allow us to determine the actual neural generators of ERPs, differences in polarity, morphology, and/or scalp distribution between Critical Words would suggest that the neural processes underlying the two ERPs are in fact different.

The following chapters report on a late sustained positive ERP component (684-1000 ms post-critical word onset) that only occurred for participants in the Untitled Discourse group, but not the Titled Discourse group. This ERP, which I classify as a "Late Sustained Frontal Positivity" (SFP), was characterized by a positive deflection, sustained morphology with no definite peak, and a frontocentral maximum scalp distribution. These characteristics disassociate the SFP from other ERPs recorded in this study. For example, the P600 was characterized by a positive polarity, "peak" morphology at approximately 556 ms and a localized scalp distribution at centroparietal electrode clusters. Because the SFP and P600 differed in morphology and scalp distribution, my conclusion will be that the neural and cognitive processes underlying these two components are to some extent different.

4.2. Temporal principal components analysis results

The temporal principal components analysis resulted in a five-factor solution that characterized 79.28% of the 1000 ms post-stimulus period. Factor 1 (27.55% of the variance) represented the Late Sustained Frontal Positivity from 684-1000 ms and peaked at 932 ms. Factor 2 (19.16%) reached its maximum during the P600 time window from 456-700 ms (peak = 556 ms). Factor 3 (16.74%) encompassed the expected N400 time window from 272-452 ms (peak = 344 ms). The variance for Factor 4 (10.49%) was maximal between 152-260 ms (peak = 180 ms) and Factor 5 (5.35%) was maximal between 104-140 ms (peak = 128 ms). Because the study's hypotheses concerned the N400 and P600 ERPs, mean amplitudes for the N400 and P600 components were extracted from the tPCA-derived temporal windows. The late temporal window (characterized the SFP) for Factor 1 was unexpected, but further explored given the large percentage of variance for which it accounted; mean amplitudes were also extracted from this time window. ERPs were averaged for each Group and Critical Word at the 15 a priori electrode clusters corresponding to the 10-10 system (Figure 3.2). Figure 4.1 shows the ERPs recorded to the three Critical Words at each Electrode Cluster for the Titled Discourse group. The same plots for the Untitled Discourse group are demonstrated in Figure 4.2.

Visual inspection of Figure 4.1 and Figure 4.2 suggest different scalp distributions during the Factor 1 time window for the Titled Discourse and Untitled Discourse groups. Expected N400 effects were recorded for both groups which appear largest at central and centroparietal clusters and become less negative from Critical Word 1 to Critical Words 2 and 3. A P600 peak followed the N400 effects and occurred at centroparietal and parietal electrode clusters for both groups and follows a similar step-wise increase in amplitude as the N400. However, the amplitudes of the P600 appear to be larger for the Titled Discourse group than the Untitled Discourse group. A large late frontal positivity (SFP) occurred for the Untitled Discourse group, but not the Titled Discourse group, at frontal

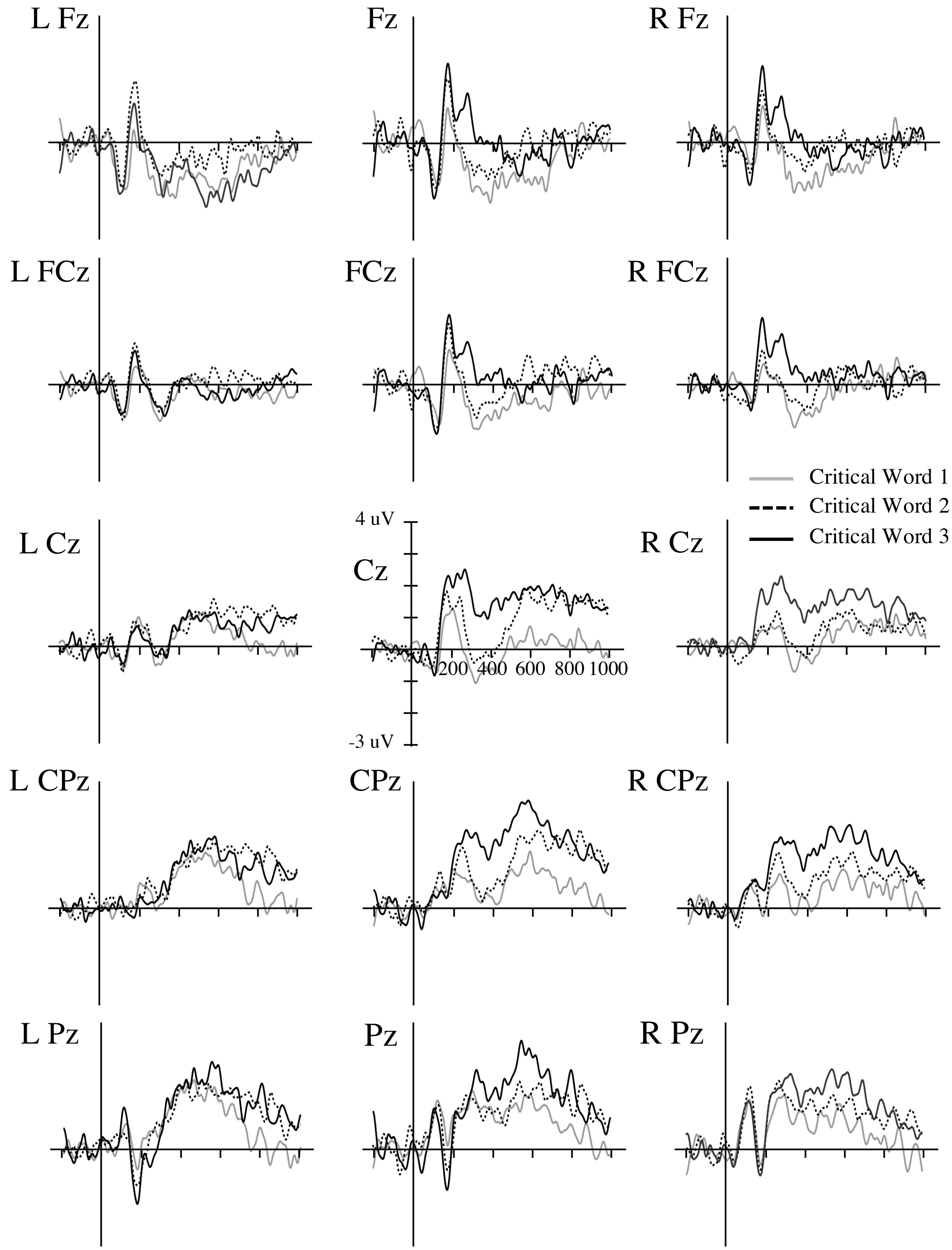

Figure 4.1. Averaged ERPs for the Titled Discourse Group at each Cluster

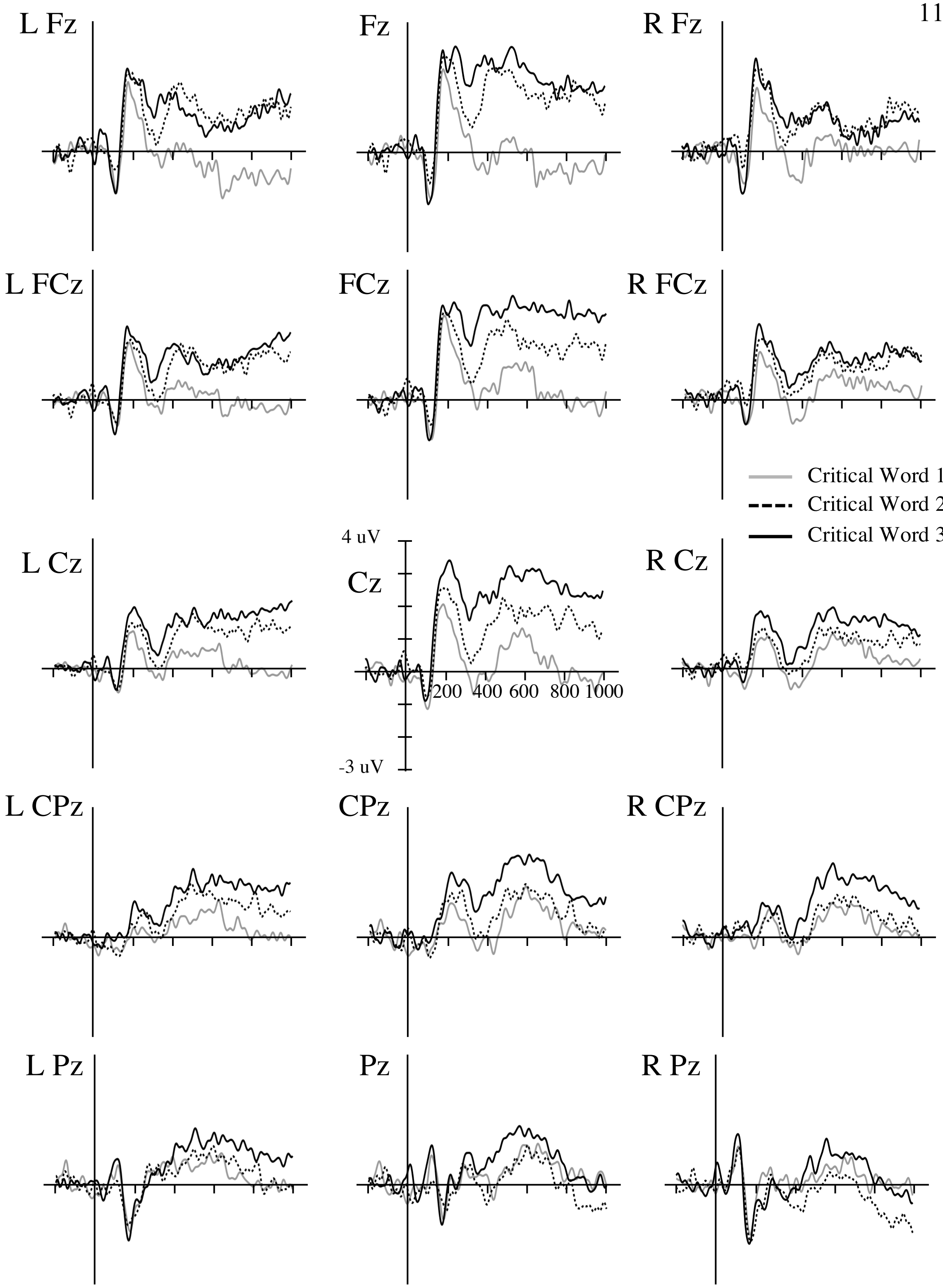

Figure 4.2. Averaged ERPs for the Untitled Discourse Group at each Cluster

and frontocentral electrode clusters. The amplitude of the SFP also increases in a step-wise pattern for the Untitled Discourse group. Systematic changes of the N400, P600, and SFP ERPs are reviewed in the following sections.

4.3. Hypothesis 1: N400 Results

To examine Hypothesis 1, a General Linear Model was conducted with N400 mean amplitude (272-452 ms post-critical word onset) as the outcome. The time window, morphology (i.e., negative deflection) and scalp distribution for the N400 recorded in this study was similar to other previously reported N400 effects (Kutas & Hillyard, 1980, 1984). Repeated measures variables included Critical Word (Critical Word 1, Critical Word 2, Critical Word 3), Laterality (left, midline, right) and Electrode Cluster (Fz, FCz, Cz, CPz, Pz).[1] The model also included a between-subjects factor of Group (Titled Discourse, Untitled Discourse). The four-factor model resulted in a significant main effect of Critical Word, $F(1.93, 67.55) = 48.72$, $p < .001$, $\eta_p^2 = .582$, which was superseded by significant two-way interactions between Critical Word and Cluster, $F(2.44, 85.32) = 3.90$, $p = .017$, $\eta_p^2 = .100$, and Critical Word and Laterality, $F(2.34, 81.72) = 5.72$, $p = .003$, $\eta_p^2 = .140$. N400 amplitudes significantly varied between Clusters at Critical Word 1, $F(4, 32) = 3.11$, $p = .029$. Specifically, N400 amplitudes were significantly larger at Pz than Cz ($p = .023$), CPz ($p = .020$), and Fz ($p = .046$).

[1] Multivariate effects are reported for significant interactions with corresponding F-statistics, p-values, and effect size estimates (η_p^2). Follow-up pairwise comparisons were performed in multivariate space in order to avoid inflation of Type I error rate (Manly, 2004, pp. 41-46). Thus separate F-statistics for pairwise comparisons were not computed in univariate space. Descriptive statistics and p-values for significant pairwise comparisons are reported.

Table 4.1

Multivariate Effects of Critical Word within each Group and Cluster

ERP	Group	Cluster	F(2, 34)	p	η_p^2
N400	Untitled	CPz	6.97	.003**	0.291
		Cz	15.83	.001**	0.482
		FCz	14.51	< .001***	.460
		Fz	10.98	< .001***	.392
	Titled	CPz	11.61	< .001***	.406
		Cz	18.10	< .001***	.516
		FCz	7.87	.002**	.317
		Fz	3.93	.029*	.188
P600	Untitled	CPz	7.63	.002**	.310
		Cz	12.54	< .001***	.425
		FCz	8.75	.001**	.340
		Fz	9.17	.001**	.351
		Pz	3.96	.029*	.189
	Titled	CPz	6.52	.004**	.277
		Cz	4.99	.012*	.227
		Fz	3.88	.030*	.186
		Pz	4.28	.022*	.201
SFP	Untitled	CPz	9.95	< .001***	.369
		Cz	19.45	< .001***	.534
		FCz	11.65	< .001***	.407
		Fz	9.93	< .001***	.369
		Pz	3.93	.029*	.188
	Titled	CPz	6.37	.004**	.272
		Cz	4.64	.017*	.214

Note. *p < .05, **p < .01, ***p < .001

N400 amplitudes did not vary between laterality sites at Pz for either Critical Word 1 (ps > .05). Table 4.1 reports multivariate effects of Critical Word within each Group and Electrode Cluster for amplitudes of the N400, P600, and Late Sustained Frontal Positivity.

Pairwise comparisons will be discussed separately, first for the Untitled Discourse group and then the Titled Discourse group. N400 amplitudes in the Untitled Discourse group significantly varied between Critical Words at CPz, Cz, FCz, and Fz. As shown in Figure 4.3, N400 amplitudes were also more negative to Critical Word 2 than Critical Word 3 at midline CPz, Cz, and FCz.

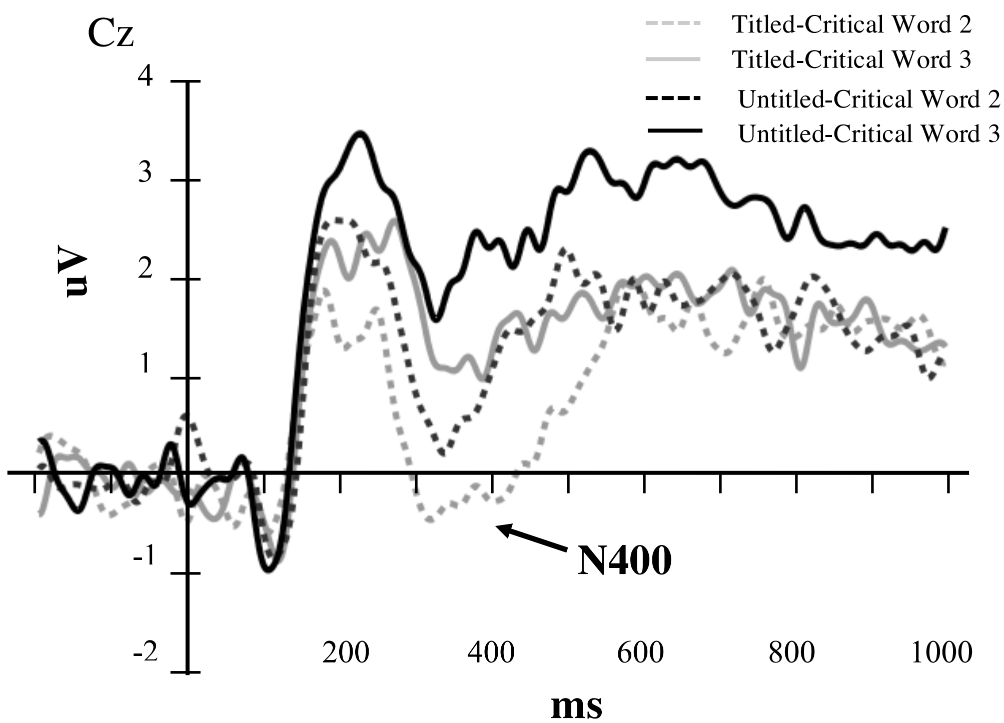

Figure 4.3. N400 effect at Midline CPz for each Group at Critical Words 2 and 3

Table 4.2 reports the significant pairwise comparisons between Critical Words on the N400 amplitude within each Cluster, Laterality, and Group. For the Untitled Discourse group, N400 amplitudes were significantly more negative to Critical Word 1 than Critical Word 2 at midline Cz, FCz, and Fz and left Fz. Larger N400 amplitudes to Critical Word 2 than Critical Word 3 occurred at midline clusters (CPz, Cz, FCz).

Table 4.2

N400 Amplitude Differences within Group, Cluster, and Laterality by Critical Word

			Mean ±SD			
Group	Cluster	Hemi	Word 1	Word 2	Word 3	p
Titled	CPz	Midline	0.45 ± 1.90		2.15 ± 1.80	.001**
		Midline		0.76 ± 1.45	2.15 ± 1.80	< .001**
		Right	0.26 ± 1.41		1.90 ± 1.28	< .001***
		Right		0.80 ± 1.18	1.90 ± 1.28	< .001***
	Cz	Midline	-0.76 ± 2.50		1.48 ± 2.70	< .001***
		Midline		-0.19 ± 2.18	1.48 ± 2.70	< .001***
		Right	-0.34 ± 1.29		1.37 ± 1.51	< .001***
		Right		0.04 ± 1.55	1.37 ± 1.51	< .001***
	FCz	Midline	-1.14 ± 3.36		0.36 ± 3.08	.010*
		Midline		-0.68 ± 3.37	0.36 ± 3.08	.005**
		Right	-0.91 ± 1.65		0.50 ± 1.92	.001**
		Right		-0.50 ± 2.11	0.50 ± 1.92	< .001**
	Fz	Midline	-1.55 ± 3.38		0.15 ± 3.74	.012*
		Midline		-0.99 ± 4.18	0.15 ± 3.74	.018*
		Right	-1.26 ± 1.99		0.14 ± 2.39	.001**
Untitled	CPz	Midline	-0.07 ± 1.81		1.37 ± 1.52	.002**
		Midline		0.42 ± 1.39	1.37 ± 1.52	.016*
	Cz	Left	0.09 ± 1.46		1.08 ± 1.34	.015*
		Midline	-0.11 ± 2.31	0.95 ± 2.10		.028*
		Midline	-0.11 ± 2.31		2.24 ± 2.38	< .001***
		Midline		0.95 ± 2.10	2.24 ± 2.38	.001**
		Right	-0.27 ± 1.03		0.63 ± 1.34	.027*
	FCz	Left	0.07 ± 1.60		1.25 ± 1.55	.010*
		Midline	0.30 ± 2.12	1.33 ± 2.88		.041*
		Midline	0.30 ± 2.12		2.48 ± 2.85	< .001***
		Midline		1.33 ± 2.88	2.48 ± 2.85	.001**
		Right	-0.37 ± 0.99		0.72 ± 1.83	.007**
	Fz	Left	-0.08 ± 2.18	1.16 ± 2.62		.021*
		Left	-0.08 ± 2.18		1.53 ± 2.44	.009**
		Midline	-0.16 ± 2.89	1.64 ± 3.60		.001**
		Midline	-0.16 ± 2.89		2.58 ± 3.44	< .001***
		Right	-0.41 ± 1.66		1.01 ± 2.33	.001**

Note. *p < .05, **p < .01, ***p < .001

Thus, the size of N400 amplitudes for the Untitled Discourse group followed a step-wise pattern, with increasingly smaller amplitudes (less negative) from Critical Word 1, to Critical Word 2 and Critical Word 3.

N400 amplitudes for the Titled Discourse group significantly differed between Critical Words at CPz, Cz, FCz, and Fz (Table 4.1). Like the Untitled Discourse group, N400 amplitudes were significantly more negative to Critical Word 1 than Critical Word 3. This effect was recorded at midline clusters (CPz, Cz, FCz, Fz), and right lateralized clusters (CPz, Cz, FCz, Fz). Larger amplitudes (more negative) to Critical Word 2 than Critical Word 3 were recorded at CPz (midline, right), Cz (midline, right), FCz (midline, right), and midline Fz. However, in contrast to the Untitled Discourse group, N400 amplitudes did not differ for Titled Discourse group between Critical Word 1 and Critical Word 2 at either cluster.

In regards to between-groups differences, the Untitled Discourse group generated significantly more negative N400 amplitudes than the Titled Discourse group to Critical Word 3 at CPz, $F(1, 35) = 4.34$, $p = .043$, $\eta_p^2 = .112$. The Untitled Discourse group generated less negative amplitudes than the Titled Discourse group to Critical Word 3 at FCz, $F(1, 35) = 4.14$, $p = .05$, $\eta_p^2 = .106$, and Fz, $F(1, 35) = 4.57$, $p = .040$, $\eta_p^2 = .116$.

4.4. Hypothesis 2: P600 Results

To examine Hypothesis 2, a GLM was performed on the P600 amplitude (456-700 ms post-critical word onset). The timing, morphology, and scalp distribution of the P600 was similar to previously reported P600 effects occurring during discourse comprehension (e.g., Hoeks et al., 2004; Nieuwland & Van Berkum, 2005). The model included the same predictors as Hypothesis 1: Critical Word, Laterality, Cluster, and Group. Significant main effects were identified for Critical Word, $F(1.72, 60.31) = 30.87$, $p < .001$, $\eta_p^2 = .469$, and Laterality, $F(1.91, 66.82) = 3.68$, $p = .032$, $\eta_p^2 = .095$. These main effects were superseded by a significant two-way interaction between Critical Word and Laterality, $F(2.79, 97.46) = 5.06$, $p = .003$, $\eta_p^2 = .126$. A significant main effect of Cluster, $F(1.28, 44.82) = 8.18$, $p = .004$, $\eta_p^2 = .189$, was superseded by a significant interaction between Cluster and Group, $F(1.28, 44.82) = 6.79$, $p = .008$, $\eta_p^2 = .162$.

Differences in P600 amplitudes by Critical Word within each Cluster and Laterality will be discussed for each group separately. As shown in Table 4.1, for the Untitled Discourse group, P600 amplitudes significantly varied between Critical Words at CPz, Cz, FCz, Fz, and Pz. P600 amplitudes for the Untitled Discourse group were not significantly different between Clusters (ps > .05). Table 4.3 demonstrates the significant pairwise comparisons for the P600 amplitude by Critical Word for each Group, Laterality, and Cluster.

For the Untitled Discourse group, P600 amplitudes were significantly larger to Critical Word 2 than Critical Word 1 at frontocentral electrode sites only (midline and left Cz, FCz, and Fz). As further discussed in section 4.5, frontally distributed effects in the P600 time window effects are interpreted as early latency variation in the Late Sustained Frontal Positivity (SFP) ERP, not the centroparietally distributed P600 ERP. P600 amplitudes were also larger to Critical Word 3 than Critical Word 1 at centroparietal clusters (CPz: left, midline, right; Cz: left, midline) and frontocentral

Table 4.3

P600 Amplitude Differences within Group, Cluster, and Laterality by Critical Word

			Mean ±SD			
Group	Cluster	Hemi	Word 1	Word 2	Word 3	p
Titled	CPz	Midline	1.27 ± 2.29		3.01 ± 2.30	.004**
		Midline		2.01 ± 1.89	3.01 ± 2.30	.008**
		Right	0.91 ± 1.72		2.37 ± 1.91	.005**
		Right		1.42 ± 1.72	2.37 ± 1.91	.009**
	Cz	Midline	0.28 ± 2.51	1.36 ± 2.63		.014*
		Midline	0.28 ± 2.51		1.89 ± 2.81	.002**
		Right	0.54 ± 1.63		1.60 ± 1.95	.013*
		Right		0.86 ± 2.05	1.60 ± 1.95	.022*
	Fz	Left		-0.43 ± 2.44	-1.57 ± 2.57	.022*
	Pz	Midline		1.68 ± 2.26	2.68 ± 2.39	.046*
		Right	1.01 ± 1.40		2.19 ± 1.66	.041*
Untitled	CPz	Left	0.68 ± 0.95		1.76 ± 1.59	.009**
		Midline	1.03 ± 1.28		2.42 ± 2.29	.004**
		Midline		1.37 ± 1.45	2.42 ± 2.29	.008**
		Right	0.78 ± 1.09		1.83 ± 1.65	.048*
		Right		1.03 ± 1.59	1.83 ± 1.65	.028*
	Cz	Left	0.40 ± 1.05	1.29 ± 1.45		.035*
		Left	0.40 ± 1.05		1.58 ± 1.55	.001**
		Midline	0.85 ± 1.65	1.88 ± 2.11		.016*
		Midline	0.85 ± 1.65		3.06 ± 2.54	< .001***
		Midline		1.88 ± 2.11	3.06 ± 2.54	.001**
	FCz	Left	0.09 ± 1.28	1.19 ± 1.48		.008**
		Left	0.09 ± 1.28		1.11 ± 1.55	.014*
		Midline	0.77 ± 1.63	1.93 ± 2.66		.015*
		Midline	0.77 ± 1.63		2.92 ± 2.96	.001**
		Midline		1.93 ± 2.66	2.92 ± 2.96	.042*
	Fz	Left	-0.58 ± 2.07	1.25 ± 2.22		.002**
		Left	-0.58 ± 2.07		0.84 ± 2.14	.035*
		Midline	-0.12 ± 2.61	2.07 ± 3.04		< .001***
		Midline	-0.12 ± 2.61		2.67 ± 3.43	.001**
	Pz	Right		0.11 ± 1.71	0.92 ± 1.76	.038*

Note. *p < .05, **p < .01, ***p < .001

clusters (FCz : left, midline; Fz: left, midline). Similarly, P600 amplitudes were larger to Critical Word 3 than Critical Word 2 at centroparietal clusters, including midline CPz and Cz (Figure 4.4) and right Cz and Pz, and frontocentral clusters (midline FCz). Thus, the

Untitled Discourse group failed to generate a centroparietal P600 effect until Critical Word 3. Unexpectedly, larger amplitudes in the P600 latency window were observed at frontocentral clusters to Critical Word 2 than Critical Word 1. Following, ERP amplitudes were larger to Critical Word 3 than Critical Word 2 at centroparietal (P600 effect) and frontocentral (SFP effect) channels.

P600 amplitudes for the Titled Discourse group significantly varied between Critical Words at CPz, Cz, Fz, and Pz, but not FCz (see Table 4.1). In contrast to the Untitled group, P600 amplitudes were significantly different between Clusters, $F(2, 34) = 9.25$, $p < .001$. Specifically, P600 amplitudes for the Titled Discourse group were significantly greater at CPz and Pz than FCz (CPz: $p < .001$; Pz: $p = .031$) and Fz (CPz: p

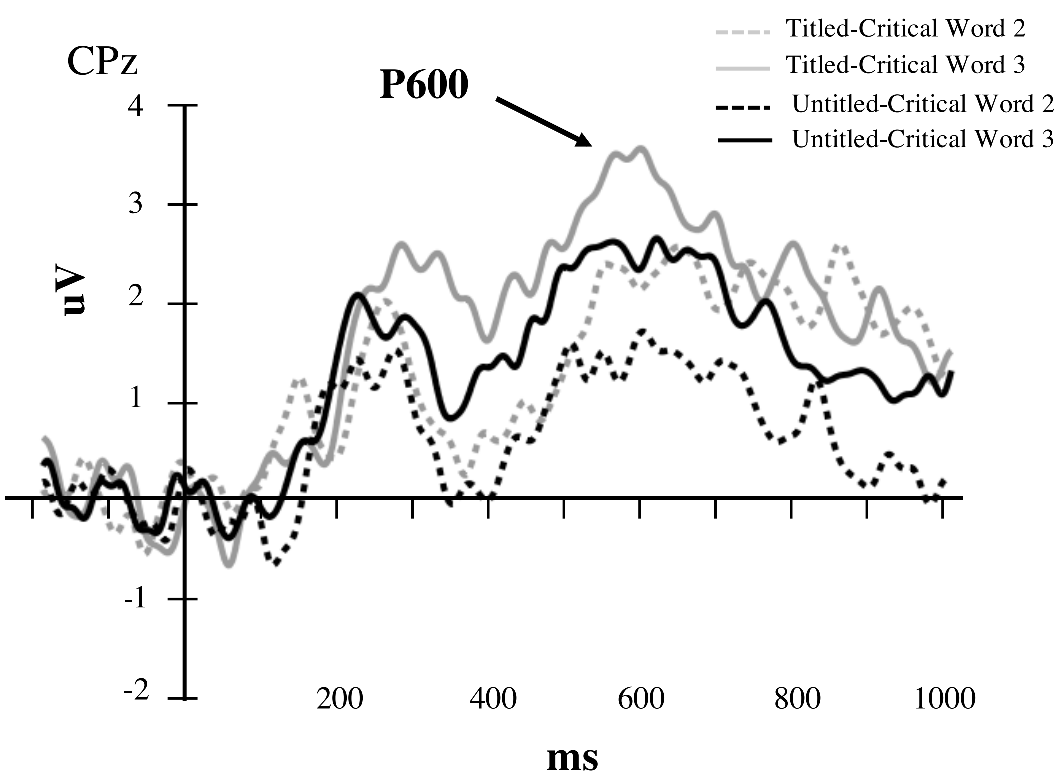

Figure 4.4. P600 effect at Midline CPz for each Group at Critical Words 2 and 3

$< .001$; Pz: $p = .014$). P600 amplitudes were not significantly different between CPz and Pz ($p > .05$). As shown in Table 4.3, amplitudes were significantly larger to Critical Word 2 than Critical Word 1 at midline Cz for the Titled Discourse group. P600 amplitudes were larger to Critical Word 3 than Critical Word 1 at CPz (midline, right) Cz (midline, right) and right Pz. As shown in Figure 4.4, P600 amplitudes were also larger to Critical Word 3 than Critical Word 2 at CPz (midline, right), right Cz, left Fz, and midline Pz.

In summary, P600 amplitudes between Critical Words followed a stepwise pattern for the Titled Discourse group: Amplitudes significantly increased from Critical Word 1 to Critical Word 2 and to Critical Word 3. However, a centroparietal P600 effect to Critical Word 2 did not occur for the Untitled Discourse group. This group of participants generated a larger P600 amplitude to Critical Word 3 than Critical Word 2. Between-groups comparisons suggest that the P600 effects followed different scalp distribution patterns for the Titled Discourse and Untitled Discourse groups. Specifically, P600 amplitudes were significantly larger for the Titled Discourse than Untitled Discourse groups to Critical Word 1 at left CPz, $F(1, 35) = 4.37$, $p = .044$, $\eta_p^2 = .11$, and left Pz $F(1, 35) = 5.49$, $p = .025$, $\eta_p^2 = 1.36$. For Critical Word 2, P600 amplitudes were also larger for the Titled Discourse group at midline Pz, $F(1, 35) = 6.09$, $p = .019$, $\eta_p^2 = .148$. However, the ERP amplitudes in the P600 time window were significantly larger for the Untitled Discourse than Titled Discourse group to Critical Word 2 at left Fz, $F(1, 35) = 4.81$, $p = .035$, $\eta_p^2 = .121$, and midline Fz, $F(1, 35) = 4.41$, $p = .043$, $\eta_p^2 = .112$. To Critical Word 3, the same pattern of results occurred at left FCz, $F(1, 35) = 7.79$, $p = .008$, $\eta_p^2 = .182$, midline FCz, $F(1, 35) = 11.18$, $p = .002$, $\eta_p^2 = .242$, left Fz, $F(1, 35) = 9.67$, $p = .004$, $\eta_p^2 = .217$, and midline Fz, $F(1, 35) = 9.51$, $p = .004$, $\eta_p^2 = .214$. Thus, these larger frontocentral amplitudes for the Untitled Discourse group are interpreted as early latency effects of the SFP component (as further discussed in the next section). In contrast, P600 effects in the Titled Discourse group were localized at the expected centroparietal and parietal clusters.

These findings suggest that both groups engaged in a similar cognitive process, as represented by the within-subjects P600 effects between Critical Words at centroparietal clusters. However, the Untitled Discourse group failed to engage in this process until Critical Word 3. The additional frontal and frontocentral effects between Critical Words for the Untitled Discourse group suggests that participants in this group may have employed an additional type of cognitive processing. These anterior-posterior differences in P600 scalp distributions between groups were further explored for the Late Sustained Frontal Positivity ERP.

4.5. Hypothesis 2: Late Sustained Frontal Positivity (SFP) Results

As previously reviewed, the temporal PCA characterized a Factor between 684-1000 ms post-Critical Word onset. This Factor constituted the largest portion of variance accounted for by the five-factor model: 27.55%. Consequently, systematic variance of this ERP (Late Sustained Frontal Positivity; "SFP") was explored using the same GLM as the N400 and P600 ERPs with mean amplitude of the SFP as the outcome. A significant main effect was identified for Critical Word, $F(1.83, 64.08) = 33.57$, $p < .001$, $\eta_p^2 = .490$, which was superseded by a two-way interaction with Group, $F(1.83, 64.08) = 4.89$, $p = .013$, $\eta_p^2 = .123$, and two-way interaction with Laterality, $F(4, 92.94) = 3.08$, $p = .037$, $\eta_p^2 = .081$. Significant three-way interactions were identified between Critical Word, Cluster, and Group, $F(2.26, 79.15) = 3.76$, $p = .023$, $\eta_p^2 = .097$, and between Critical Word, Laterality, and Cluster, $F(6.45, 225.79) = 2.23$, $p = .037$, $\eta_p^2 = .060$.

SFP amplitudes for the Untitled Discourse group varied between Critical Words and were distributed at all electrode sites (see Table 4.1). As shown in Table 4.4, SFP amplitudes for the Untitled Discourse group were significantly larger to Critical Word 2 than Critical Word 1 at left and midline Cz, FCz, and Fz. These effects are similar to the P600 latency effects between Critical Words 1 and 2. This provides further evidence that the frontocentral effects in the P600 latency window for this group reflect early latency variation in the SFP not the P600. SFP amplitude differences between Critical Word 3 and Critical Word 1 occurred at CPz (left, right), Cz (left, midline, right), FCz (left, midline), and Fz (left, midline). SFP amplitudes to Critical Word 3 than Critical Word 2 were significantly larger for the Untitled Discourse group at right CPz, midline Cz, and midline FCz. Such centroparietal SFP effects at Critical Word 3 are interpreted as late variation in the P600 ERP. The centroparietal P600 effects to Critical Word 3 for the Untitled Discourse group support this point. Visual inspection of Figure 4.2 suggests that the morphology and scalp distribution of the SFP effects (no definite peak, frontocentral distribution) are distinct from the P600 ERP (clear peak, centroparietal distributed).

Pairwise comparisons for the Titled Discourse group between Critical Words within each Cluster and Laterality are represented in Table 4.4. Larger amplitudes

Table 4.4. Late Sustained Frontal Positivity Amplitude Differences within Group, Cluster, and Laterality by Critical Word

			Mean ± SD			
Group	Cluster	Hemi	Word 1	Word 2	Word 3	p
Titled	CPz	Left	0.39 ± 1.71	1.60 ± 1.72		.020*
		Midline	0.59 ± 2.15	1.99 ± 2.16		.008**
		Midline	0.59 ± 2.15		1.98 ± 1.83	.016*
	Cz	Midline	0.21 ± 2.27	1.60 ± 2.24		.005**
		Midline	0.21 ± 2.27		1.68 ± 1.85	.004**
	Pz	Left	0.12 ± 2.07	1.23 ± 1.87		.047*
Untitled	CPz	Left	0.13 ± 1.09		1.61 ± 1.16	.001**
		Right	0.28 ± 0.98		1.45 ± 1.32	.016*
		Right		0.57 ± 1.13	1.45 ± 1.32	.020*
	Cz	Left	-0.10 ± 1.12	1.23 ± 1.54		.007**
		Left	-0.10 ± 1.12		1.69 ± 1.40	< .001***
		Midline	-0.07 ± 1.41	1.61 ± 1.69		.001**
		Midline	-0.07 ± 1.41		2.57 ± 1.81	< .001***
		Midline		1.61 ± 1.69	2.57 ± 1.81	.014*
		Right	0.35 ± 0.91		1.46 ± 1.21	.022*
	FCz	Left	-0.25 ± 1.12	1.29 ± 1.65		.001**
		Left	-0.25 ± 1.12		1.60 ± 2.00	< .001***
		Midline	-0.72 ± 1.45	1.65 ± 2.37		.002**
		Midline	-0.72 ± 1.45		2.73 ± 2.67	< .001***
		Midline		1.65 ± 2.37	2.73 ± 2.67	.044*
	Fz	Left	-0.84 ± 1.62	1.28 ± 2.26		.001**
		Left	-0.84 ± 1.62		1.22 ± 2.66	.006**
		Midline	-0.62 ± 2.22	1.74 ± 2.98		< .001***
		Midline	-0.62 ± 2.22		2.01 ± 3.29	.004**
	Pz	Right		-0.75 ± 1.70	0.19 ± 1.64	.048*

Note. *p < .05, **p < .01, ***p < .001

occurred for Critical Word 2 than Critical Word 1 at left and midline CPz, midline Cz, and left Pz. This group also generated larger amplitudes to Critical Word 3 than Critical Word 1 at midline CPz and midline Cz. These centroparietal effects are interpreted as late variation in the P600 ERP. Unlike the Untitled Discourse group, the Titled Discourse group did not generate significant amplitude differences between Critical Word 2 and

Critical Word 3 in this time window. The lack of amplitude changes at frontocentral/frontal clusters suggests that the Titled Discourse group did not engage in the neurocognitive mechanisms that the SFP reflects.

As shown in Figure 4.5, between-groups comparisons indicate that the differing scalp distributions in the SFP latency window between the Untitled Discourse and Titled Discourse groups became more prevalent than during the P600 time window. Larger SFP amplitudes occurred for the Untitled Discourse group than Titled Discourse group to

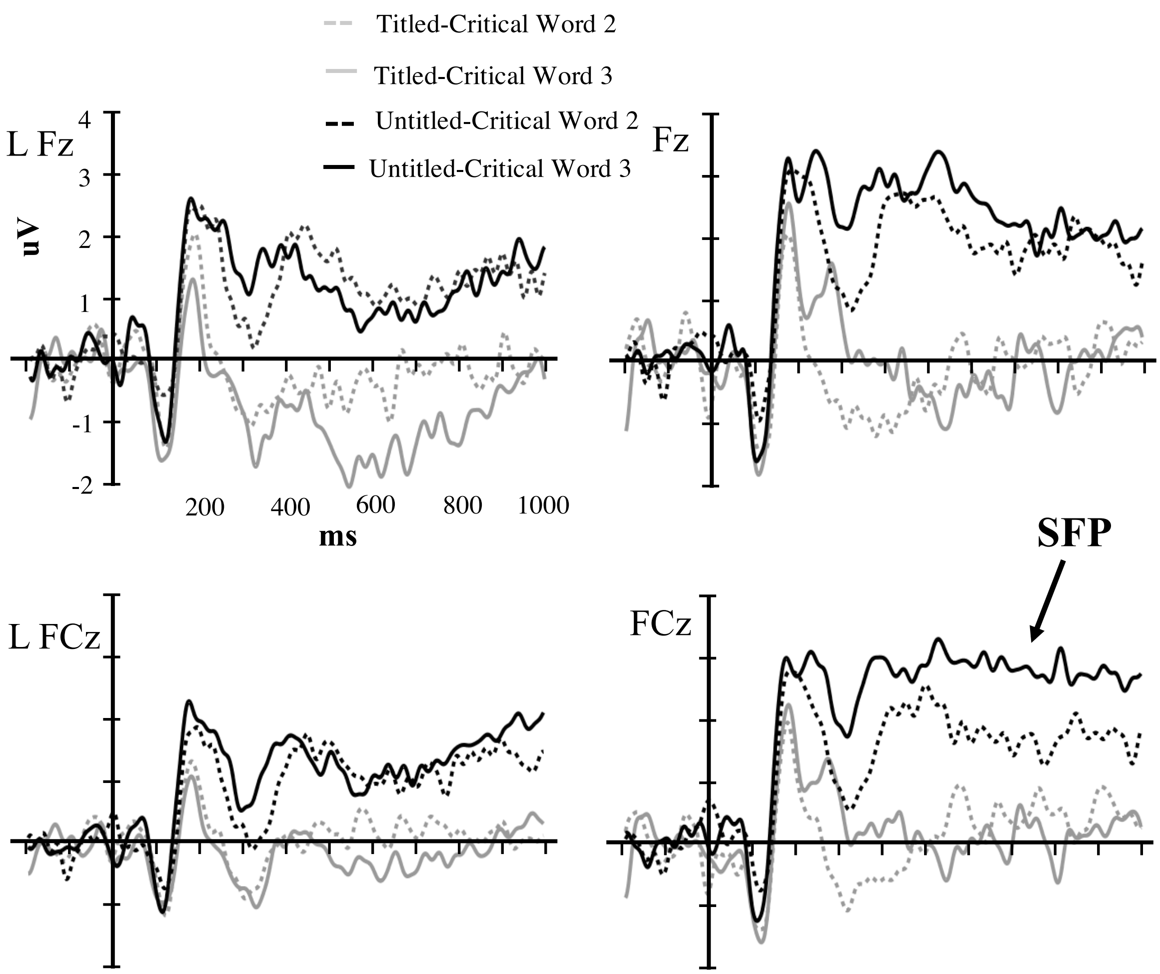

Figure 4.5. SFP effect for each Group at Left and Midline Frontal and Frontocentral Clusters to Critical Words 2 and 3

Critical Word 2 at left FCz, $F(1, 35) = 5.44$, $p = .026$, $\eta_p^2 = .134$. The Untitled Discourse group also generated larger amplitudes to Critical Word 3 than the Titled Discourse group at left Cz, $F(1, 35) = 4.39$, $p = .043$, $\eta_p^2 = .111$, left FCz $F(1, 35) = 7.04$,p $= .012$, $\eta_p^2 = .167$, left Fz, $F(1, 35) = 5.67$, $p = .023$, $\eta_p^2 = .139$, and midline FCz, $F(1, 35) = 10.25$, $p = .003$, $\eta_p^2 = .226$. However, SFP amplitudes were larger for the Titled Discourse group than the Untitled Discourse group to Critical Word 2 at midline CPz, $F(1, 35) = 5.34$, $p = .027$, $\eta_p^2 = .132$, and right Pz, $F(1, 35) = 10.50$, $p = .003$, $\eta_p^2 = .231$.

These patterns of results qualify the differences in morphology and scalp distribution of the late positive effects between the Titled Discourse and Untitled Discourse groups. Specifically, that the SFP was characterized as a sustained positivity with no definite peak with a frontocentral maximum. However, no frontal or frontocentral SFP effects occurred for the Titled Discourse group. As shown in Figures 4.1 and 4.3, the Titled group's SFP latency window effects at CPz and Pz likely reflect later-occurring variation of the peak P600 deflection. These differences in scalp distribution between the Untitled Discourse and Titled Discourse groups are further interpreted in Chapter 5.

4.6. Hypothesis 3: N400 and Confrontation Naming Abilities

To examine Hypothesis 3, accuracy on the Boston Naming Test (BNT) was added as a within-subjects quantitative covariate to the GLM from section 4.2. Each subject's BNT accuracy was mean-centered such that the mean value (50.73) was centered at zero. The modified GLM resulted in a null effect of BNT accuracy, $F(1, 34) = 0.38$, $p = .53$. BNT accuracy did not significantly interact with any combinations of other variables in the model (ps > .05). Planned comparisons were selectively examined at CPz and Cz Clusters (Kutas & Federmeier, 2000, 2011; Kutas & Hillyard, 1984) while holding constant mean-centered BNT accuracy at three different levels: low (-1 SD below mean = -4.65), mean (mean = 0), and high (+1 SD above the mean = 4.65).[2] This procedure was performed in order to examine the extent to which varying levels of BNT accuracy influenced N400 amplitudes.

Table 4.5 demonstrates the significant multivariate effects of Critical Word within each Group and Cluster on the N400 when BNT Accuracy was low, mean and high.

Table 4.5

Multivariate Effects of Critical Word within each Group and Cluster on the N400 when BNT Accuracy at Three Different Levels

BNT Accuracy	Group	Cluster	$F(2, 33)$	p	η_p^2
Low	Untitled	Cz	6.74	.004**	.290
	Titled	CPz	11.60	.004**	.281
		Cz	10.78	< .001***	.395
Mean	Untitled	CPz	6.49	.004**	.282
		Cz	15.04	< .001***	.477
	Titled	CPz	12.05	< .001***	.422
		Cz	18.50	< .001***	.529
High	Untitled	CPz	6.88	.003**	.294
		Cz	13.94	< .001***	.458
	Titled	CPz	9.60	.001**	.368
		Cz	13.80	< .001***	.455

Note. $*p < .05$, $**p < .01$, $***p < .001$

When BNT accuracy was -1 SD below the mean, participants in the Titled Discourse group generated significantly different N400 amplitudes between Critical Words at CPz

[2] This approach is recommended for examining interactions with continuous covariates (Taylor, 2011), rather than artificially transforming continuous variables into categorical groups. The latter approach is strongly cautioned against as it leads to an underpowered, underestimate of a predictor, particularly when testing interactions (Aiken & West, 1991, pp. 167-168; MacCallum et al., 2002; Maxwell & Delaney, 1993).

and Cz. Participants in the Untitled Discourse group generated significantly different N400 amplitudes between Critical Words only at Cz. When BNT Accuracy was at the mean and +1 SD above the mean, N400 amplitudes were different between Critical Words at CPz and Cz for both groups.

Table 4.6 reports descriptive statistics for N400 amplitudes for each Critical Word within each Group and Cluster when BNT Accuracy was at the three varying levels. When BNT Accuracy was at low, mean, and high levels for the Titled Discourse group, N400 amplitudes were significantly larger to Critical Word 3 than Critical Word 2 and Critical Word 1 at CPz and Cz. The same effects occurred at CPz and Cz for the Untitled Discourse group when accuracy was at the mean and higher. However, when accuracy was low for the Untitled Discourse group, N400 amplitudes were larger to Critical Word 3 than Critical Words 2 and Critical Word 1 only at Cz.

To further explore Hypothesis 3, average response time (seconds) on the Boston Naming Test was mean-centered (1.49) and added as a within-subjects quantitative covariate to the GLM from section 4.2. Results were similar to the GLM with BNT accuracy as the within-subjects covariate: There was no significant effect of BNT response time, $F(1, 34) = 0.11$, $p = .747$, nor did response time did not significantly interact with any combinations of other variables in the model ($ps > .05$). Planned comparisons were selectively examined for the effect of BNT response time on N400 amplitudes at CPz and Cz electrode clusters while holding constant mean-centered BNT response time at three different levels: slow (+1 SD above the mean = 0.54), mean (mean = 0), and fast (-1 SD below mean = -0.54).

Table 4.6

N400 Amplitude Differences between Critical Words within each Group and Cluster when BNT Accuracy is Constant at Three Different Levels

		Mean ± SE			
Group	Cluster	Word 1	Word 2	Word 3	p
BNT Accuracy = Low					
Titled	CPz	0.63 ± .032		1.70 ± 0.32	.007**
	CPz		0.96 ± 0.47	1.70 ± 0.32	.019*
	Cz	-0.16 ± .39		1.06 ± 0.44	.001**
	Cz		0.12 ± 0.42	1.06 ± 0.44	.001**
Untitled	Cz	0.12 ± 0.42		1.33 ± 0.48	.001**
	Cz		0.70 ± 0.46	1.33 ± 0.48	.001**
BNT Accuracy= Mean					
Titled	CPz	0.49 ± 0.28		1.72 ± 0.28	< .001***
	CPz		0.82 ± 0.23	1.72 ± 0.28	.001**
	Cz	-0.35 ± 0.33		1.04 ± 0.38	< .001***
	Cz		-0.01 ± 0.37	1.04 ± 0.38	< .001***
Untitled	CPz	0.01 ± 0.27		0.92 ± 0.27	.006**
	CPz		0.32 ± 0.23	0.92 ± 0.27	.024*
	Cz	-0.07 ± 0.33		1.32 ± 0.37	< .001***
	Cz		0.57 ± 0.36	1.32 ± 0.37	.002**
BNT Accuracy = High					
Titled	CPz	0.35 ± 0.36		1.74 ± 0.36	.001**
	CPz		0.68 ± 0.30	1.74 ± 0.36	.002**
	Cz	-0.53 ± 0.43		1.03 ± 0.49	< .001***
	Cz		-0.13 ± 0.47	1.03 ± 0.49	< .001***
Untitled	CPz	-0.13 ± 0.31		0.94 ± 0.32	.006**
	CPz		0.18 ± 0.27	0.94 ± 0.32	.014*
	Cz	-0.25 ± 0.38		1.30 ± 0.43	< .001***
	Cz		0.44 ± 0.42	1.30 ± 0.43	.002**

Note. *p < .05, **p < .01, ***p < .001

Table 4.7 demonstrates the significant multivariate effects of Critical Word within each Group and Electrode Cluster on the N400 when BNT Response Time was at slow, mean and fast levels. When BNT response time was at mean and fast (- 1 SD below the mean), N400 amplitudes differed between Critical Words for both the Titled Discourse

Table 4.7

Multivariate Effects of Critical Word within each Group and Cluster on the N400 when BNT Response Time at Three Different Levels

BNT Response Time	Group	Cluster	F(2, 33)	p	η^2_p
Slow	Untitled	Cz	8.60	< .001***	.343
	Titled	CPz	3.96	.029	.193
		Cz	7.61	.002**	.316
Mean	Untitled	CPz	7.81	.002**	.321
		Cz	16.50	< .001***	.500
	Titled	CPz	11.30	< .001***	.406
		Cz	17.55	< .001***	.515
Fast	Untitled	CPz	7.92	< .001***	.324
		Cz	13.18	< .001***	.444
	Titled	CPz	12.84	< .001**	.438
		Cz	17.21	< .001***	.511

Note. *p < .05, **p < .01, ***p < .001

and Untitled Discourse groups at CPz and Cz. When BNT response time was slow, the same relationship occurred for the Titled Discourse group. However, significant between-Critical Word effects for Untitled Discourse group only occurred at Cz.

As shown in Table 4.8, a similar pattern of between-word effects emerged for the low-accuracy and low-response time groups. Specifically, when BNT response time was slow for the Untitled Discourse N400 amplitudes were larger to Critical Word 3 than Critical Word 2 and Critical Word 1 at Cz, but not CPz, clusters. However, when response time was at the mean and +1 SD above the mean for the Untitled Discourse group, N400 amplitudes were larger to Critical Word 3 than Critical Word 1 and 2 at both CPz and Cz. Thus, differences in N400 amplitudes between Critical Words were similar regardless of confrontation naming abilities. There was, however, no difference at CPz

Table 4.8

N400 Amplitude Differences between Critical Words within each Group and Cluster when BNT Response Time at Three Different Levels

Group	Cluster	Word 1	Word 2	Word 3	p
		Mean ± SE			
BNT Response Time = Slow					
Titled	CPz	0.53 ± 0.35		1.46 ± 0.35	.038*
	Cz	-0.24 ± 0.43		0.93 ± 0.45	.006**
	Cz		0.11 ± 0.47	0.93 ± 0.45	.009**
Untitled	Cz	-0.38 ± 0.39		1.23 ± 0.43	.001**
	Cz		0.63 ± 0.42	1.23± 0.43	.039*
BNT Response Time = Mean					
Titled	CPz	0.51 ± 0.28		1.69 ± 0.27	< .001***
	CPz		0.85 ± 0.23	1.69 ± 0.27	.001**
	Cz	-0.31 ± 0.34		1.03 ± 0.38	< .001***
	Cz		0.02 ± 0.37	1.03 ± 0.38	< .001**
Untitled	CPz	-0.2 ± 0.27		0.96 ± 0.26	.003**
	CPz		0.30 ± 0.02	0.96 ± 0.26	.010*
	Cz	-0.10 ± 0.33		1.33 ± 0.37	< .001***
	Cz		0.54 ± 0.36	1.33 ± 0.37	.001**
BNT Response Time = Fast					
Titled	CPz	0.50 ± 0.32		1.91 ± 0.32	< .001***
	CPz		0.83 ± 0.27	1.91 ± 0.32	< .001***
	Cz	-0.37 ± 0.39		1.14 ± 0.44	< .001***
	Cz		-0.06 ± 0.43	1.14 ± 0.44	< .001***
Untitled	CPz	-0.4 ± 0.35		1.18 ± 0.34	.004**
	CPz		0.28 ± 0.30	1.18 ± 0.34	.006**
	Cz	-0.17 ± 0.42		1.43 ± 0.48	<.001***
	Cz		0.46 ± 0.46	1.43 ± 0.48	.001**

Note. *p < .05, **p < .01, ***p < .001

between Critical Words for the Untitled Discourse group when accuracy and response time were low. This suggests that individuals with lower confrontation naming abilities may recruit different neural resources to support lexical-semantic retrieval during the comprehension of contextually ambiguous discourses.

4.7. Hypothesis 4: P600/SFP and Working Memory Capacity

To examine Hypothesis 4, Reading Span performance was centered on the sample mean (53.676) and added as a within-subjects quantitative covariate to the GLM from section 4.3. The modified GLM resulted in a null effect of Reading Span, $F(1, 34) = 0.14$, $p = .714$. Reading Span did not significantly interact with any combinations of other variables ($ps > .05$). Table 4.9 demonstrates the significant multivariate effects of Critical Word within each Group and Cluster on the P600 when Reading Span performance was held at low (-1 SD below mean = -11.87), mean (mean = 0), and high (+1 SD above the mean = 11.87) levels.

When Reading Span was -1 SD below the mean for the Titled Discourse group, P600 amplitudes were significantly different between Critical Words at Fz. For the Untitled Discourse group, however, P600 amplitudes significantly varied between Critical Words at CPz, Cz, FCz, and Fz. At the mean, P600 amplitudes significantly differed between Critical Words for both groups at CPz, Cz, Fz, and Pz. When Reading Span was +1 SD above the mean, the Titled Discourse group generated significantly different P600 amplitudes between Critical Words at CPz, Cz, and Pz, whereas differences for the Untitled Discourse group occurred at all electrode sites.

Pairwise comparisons between Critical Words for the three different levels of Reading Span performance are reported in Table 4.10. When Reading Span was -1 SD below the mean, P600 amplitudes did not differ between Critical Words in the Titled Discourse group. Otherwise, P600 amplitudes were larger to Critical Word 3 than Critical Word 1 at Cz and CPz for both groups regardless of Reading Span. P600 amplitudes were also larger to Critical Word 3 than Critical Word 2 at CPz for the

Table 4.9

Multivariate Effects of Critical Word within each Group and Cluster on the P600 when Reading Span at Three Different Levels

Reading Span	Group	Cluster	F(2, 33)	p	η^2_p
Low	Untitled	CPz	4.14	.025*	.200
		Cz	7.89	.002**	.323
		FCz	6.49	.004**	.282
		Fz	8.22	.001**	.335
	Titled	Fz	3.52	.041*	.176
Mean	Untitled	CPz	8.04	.001**	.328
		Cz	12.49	< .001***	.431
		FCz	8.26	.001**	.334
		Fz	8.38	.001**	.337
		Pz	4.11	.026*	.199
	Titled	CPz	5.82	.007**	.261
		Cz	4.45	.018*	.216
		Fz	4.14	.025*	.200
		Pz	3.72	.035*	.184
High	Untitled	CPz	6.47	.004**	.282
		Cz	8.52	.001**	.341
		FCz	4.61	.017*	.218
		Fz	3.52	.041*	.176
		Pz	3.71	.035*	.184
	Titled	CPz	6.82	.003**	.292
		Cz	4.44	.020*	.212
		Pz	5.15	.011*	.238

Note. *p < .05, **p < .01, ***p < .001

Untitled Discourse and Titled Discourse groups across Reading Span performance (except Titled Discourse group at low levels). Regardless of Reading Span performance for the Untitled Discourse group, the centroparietal P600 did not vary between Critical Words 1 and 2. Larger P600 amplitudes occurred to Critical Word 3 than Critical Word 2 irrespective of Reading Span. Taken together these effects may indicate that working memory capacity does not influence context-updating processing (P600; Brouwer et al., 2012) when the context of a discourse is ambiguous.

Table 4.10

P600 Amplitude Differences between Critical Words within each Group and Cluster when Reading Span Performance at Three Different Levels

		Mean ± SE			
Group	Cluster	Word 1	Word 2	Word 3	p
Reading Span Performance = Low					
Untitled	CPz		0.83 ± 0.34	1.76 ± 0.45	.036*
	Cz	0.76 ± 0.37		2.09 ± 0.47	.001**
	FCz	0.70 ± 0.37		2.01 ± 0.48	.004**
Reading Span Performance = Mean					
Titled	CPz	1.22 ± 0.30		2.37 ± 0.40	.008**
	CPz		1.75 ± 0.33	2.37 ± 0.40	.037*
	Cz	0.54 ± 0.33		1.46 ± 0.42	.013*
	Pz		1.56 ± 0.38	2.23 ± 0.46	.046*
Untitled	CPz	0.83 ± 0.29		2.07 ± 0.39	.003**
	CPz		1.29 ± 0.32	2.07 ± 0.39	.005**
	Cz	0.66 ± 0.32		2.08 ± 0.41	< .001***
	Cz		1.40 ± 0.41	2.08 ± 0.41	.006**
	FCz	0.41 ± 0.32		1.73 ± 0.42	< .001***
	Fz	-0.25 ± 0.50	1.33 ± 0.60		.001**
	Pz		0.65 ± 0.37	1.39 ± 0.45	.021*
Reading Span Performance = High					
Titled	CPz	1.22 ± 0.34		2.68 ± 0.45	.003**
	CPz		1.98 ± 0.37	2.68 ± 0.45	.036*
	Cz	0.44 ± 0.37		1.45 ± 0.48	.017*
	Pz	1.38 ± 0.33		2.80 ± 0.52	.015*
			2.02 ± 0.43	2.80 ± 0.52	.041*
Untitled	CPz	0.84 ± 0.39		2.38 ± 0.52	.006**
	CPz		1.51 ± 0.42	2.38 ± 0.52	.021*
	Cz	0.56 ± 0.42		2.08 ± 0.54	< .001***
	Cz		1.25 ± 0.55	2.08 ± 0.54	.014*
	FCz	0.12 ± 0.43	0.95 ± 0.62		.013*
	Fz	0.72 ± 0.37	1.10 ± 0.48		.038*

Note. *p < .05, **p < .01, ***p < .001

As discussed in section 4.4, the Untitled Discourse group also generated a Late Sustained Frontal Late Positivity ERP between Critical Words. For this reason, the same

GLM model from section 4.4 was applied to the amplitude of the SFP (684-1000 ms) with mean-centered Reading Span performance as a within-subjects covariate. Reading Span was not a significant predictor in the model, $F(1, 34) = 0.00$, $p = .949$. Planned comparisons examined the extent to which the effects of Critical Word within each Cluster and Group varied for the same three levels of mean-centered Reading Span as previously described (low = -11.87; mean = 0; high = 11.87). Table 4.11 lists the significant multivariate effects of Critical Word within each Group and Cluster on the late positive ERP when Reading Span performance was at these three levels.

Table 4.11

Multivariate Effects of Critical Word within each Group and Cluster on the Late Positive ERP when Reading Span at Three Different Levels

Reading Span	Group	Cluster	F(2, 33)	p	η^2_p
Low	Untitled	CPz	8.19	.001**	.332
		Cz	15.64	< .001***	.487
		FCz	8.26	.001**	.333
		Fz	9.68	< .001***	.370
		Pz	4.65	.017*	.220
	Titled	CPz	3.62	.038*	.180
		Cz	3.50	.042*	.172
Mean	Untitled	CPz	9.20	.001**	.358
		Cz	18.12	< .001***	.523
		FCz	11.14	< .001***	.403
		Fz	9.06	.001**	.355
		Pz	3.41	.045	.171
	Titled	CPz	6.05	.006**	.268
		Cz	4.70	.016*	.222
High	Untitled	CPz	4.44	.020*	.212
		Cz	8.97	.001**	.352
		FCz	6.69	.004**	.289
		Fz	3.84	.032*	.189
	Titled	CPz	4.69	.016*	.221

Note. *p < .05, **p < .01, ***p < .001

For all levels of Reading Span, SFP amplitudes were significantly larger to Critical Word 2 than Critical Word 1 at Cz, FCz, and Fz (Table 4.12). However, when Reading Span was at the mean or -1 SD below the mean for the Untitled Discourse group, SFP amplitudes were significantly larger to Critical Word 3 than Critical Word 2 at centroparietal clusters and central clusters (CPz and Cz). However, when Reading Span was high, SFP amplitudes did not differ between Critical Word 3 and Critical Word 2 at any cluster for the Untitled group. Taken together, these results may indicate that working memory capacity does not moderate the frontocentral SFP amplitudes occurring early in the contextually ambiguous discourses (Critical Word 2). Irrespective of Reading Span, frontocentral SFP amplitudes did not significantly vary between Critical Words 2 and 3, although larger SFP amplitudes were noted and central clusters. When Reading Span was at the mean or below, late-latency centroparietal P600 amplitudes were also larger to Critical Word 3 than Critical Word 2. Taken together, these results suggest that working memory capacity has little effect on the amplitudes of the P600 or SFP ERP components. When Reading Span is high, however, centroparietal P600 effects to Critical Word 3 may follow a more “peak-like” morphology (no P600 centroparietal effect during SFP latency window), than when Reading Span is low or at the mean (P600 centroparietal effect during SFP latency window).

Table 4.12

Late Positivity Amplitude Differences between Critical Words within each Group and Cluster when Reading Span at Three Different Levels

		Mean ± SE			
Group	Cluster	Word 1	Word 2	Word 3	p
Reading Span Performance = Low					
Titled	CPz	0.28 ± 0.37		1.44 ± 0.38	.032*
	Cz	0.17 ± 0.41		1.25 ± 0.42	.036*
Untitled	CPz	0.11 ± 0.31		1.39 ± 0.33	.004**
	CPz		0.51 ± 0.35	1.39 ± 0.33	.010*
	Cz	-0.02 ± 0.35	1.25 ± 0.40		.003**
	Cz	-0.02 ± 0.35		1.93 ± 0.35	< .001***
	Cz		1.25 ± 0.40	1.93 ± 0.35	.027*
	FCz	0.18 ± 0.38	1.62 ± 0.47		.002**
	FCz	0.18 ± 0.38		2.01 ± 0.47	.001**
	Fz	-0.27 ± 0.57	1.87 ± 0.67		< .001***
	Fz	-0.27 ± 0.57		1.52 ± 0.65	.032*
Reading Span Performance = Mean					
Titled	CPz	0.47 ± 0.28	1.51 ± 0.31		.011*
	CPz	0.47 ± 0.28		1.57 ± 0.29	.005**
	Cz	0.27 ± 0.31	1.11 ± 0.36		.035*
	Cz	0.27 ± 0.31		1.23 ± 0.32	.013*
Untitled	CPz	0.30 ± 0.27		1.52 ± 0.28	.001**
	CPz		0.77 ± 0.31	1.52 ± 0.28	.010*
	Cz	0.08 ± 0.31	1.25 ± 0.35		.002**
	Cz	0.08 ± 0.31		1.90 ± 0.31	< .001***
	Cz		1.25 ± 0.35	1.90 ± 0.31	.012*
	FCz	-0.51 ± 0.33	1.31 ± 0.41		.001**
	FCz	-0.51 ± 0.33		1.87 ± 0.41	< .001***
	Fz	-0.54 ± 0.50	1.25 ± 0.58		< .001***
	Fz	-0.54 ± 0.50		1.32 ± 0.57	.008**
	Pz		1.25 ± 0.58	1.32 ± 0.57	.042*
Reading Span Performance = High					
Titled	CPz	0.66 ± 0.32	1.77 ± 0.36		.019*
	CPz	0.66 ± 0.32		1.71 ± 0.33	.022*
Untitled	CPz	0.49 ± 0.36		1.66 ± 0.38	.024*
	Cz	0.19 ± 0.40	1.24 ± 0.47		.040*
	Cz	0.19 ± 0.40		1.88 ± 0.41	.001**
	FCz	-0.28 ± 0.44	1.00 ± 0.54		.020*
	FCz	-0.28 ± 0.44		1.72 ± 0.55	.002**
	Fz	-0.48 ± 0.58	-0.54 ± 0.68		.043*

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

CHAPTER 5

DISCUSSION

5.1. General discussion

Two studies were performed to investigate the temporal structure and organization of psycholinguistic processing during the comprehension of contextually ambiguous discourses. In Study 1, a novel stimulus set of 25, three-sentence long discourses were developed in which their contexts were initially ambiguous but became significantly clearer when participants read the last word of the second sentence (Critical Word 2) and even more so after the last word of the third sentence (Critical Word 3). When there was no title presented before the discourses, the cloze probability (i.e., expectedness) did not vary between the last words in each of the three sentences. However, when the title was provided, cloze probability was significantly higher to Critical Word 3 than Critical Word 2 and Critical Word 1. These results established that when reading contextually ambiguous coherent discourses (Untitled Discourse group), participants updated their mental model of the discourse after reading words that were semantically related to the theme of the discourses.

In Study 2, high-density event-related potentials were recorded when participants read 25 of the discourses used in Study 1. The study followed a mixed-groups design in which participants read the discourses with or without the title presented immediately before each of the discourses. As a result, the title provided a subtle manipulation in which the contexts of the discourses were available (Titled Discourse group) or initially ambiguous (Untitled Discourse group). The ERPs of interest were the N400 and P600 recorded to Critical Words 1-3. An unexpected Late Sustained Frontal Positivity ERP Component occurred for the Untitled Discourse group and was thus further explored.

St. George and colleagues (1994) reported that N400 amplitudes were more negative to words in contextually ambiguous coherent discourses than discourses in which the contexts were known. Research also established that N400 amplitudes were positively related to the cloze probability of words (DeLong et al., 2005; Kutas & Hillyard, 1984). Based on these findings, I hypothesized that N400 amplitudes would not vary for participants in the Untitled Discourse group. Hypothesis 1 was based on findings from Study 1—cloze probabilities of Critical Words did not vary for participants in the Untitled group. Cloze probabilities were significantly greater to Critical Word 3 compared to Critical Word 1 and 2 for those in the Titled Discourse group. Thus, it was hypothesized that N400 amplitudes would be more negative to Critical Words 1 and 2 than Critical Word 3.

Hypothesis 1 was partially supported. As hypothesized, N400 results for the Titled Discourse group corresponded to the cloze probability results from Study 1: Amplitudes did not vary between Critical Words 1 and 2, but were less negative to Critical Word 3 than Critical Words 1 and 2. This latter effect was also observed for the Untitled Discourse group. Unexpectedly, N400 amplitudes were also more negative to Critical Word 1 than to Critical Word 2 only for the Untitled Discourse group at Cz, FCz, and Fz. These effects are interpreted as a more substantial ease of lexical-semantic retrieval at Critical Word 2 and Critical Word 3. A potential basis for these patterns of N400 effects becomes clearer when we consider how the amplitude of the P600 ERP changed between Critical Words.

Prior research suggested that the centroparietal P600 ERP might reflect context updating during discourse comprehension (e.g., Brouwer et al., 2012) or the establishment of a new discourse referent (e.g., Burkhardt, 2006; Kaan et al., 2007). Based on these views, it was hypothesized that P600 amplitudes would mirror the title-identification findings from Study 1. Specifically, for participants in the Untitled Discourse group, P600 amplitudes would increase from Critical Word 1 to Critical Word 2 and Critical Word 3. This would suggest that larger P600 amplitudes reflect, at least in part, participants' context updating within coherent discourse.

Hypothesis 2 was partially supported: P600 amplitudes increased from Critical Words 1 to Critical Word 2 and to Critical Word 3 for the Titled Discourse group. In the Untitled Discourse group only, an unexpected Late Sustained Frontal Positivity emerged (SFP: 684-1000 ms post-critical word onset). The SFP was characterized by a sustained frontal/frontocentral positivity. For the Untitled Discourse group, SFP amplitudes became increasingly larger from Critical Word 1 to Critical Word 2 and to Critical Word 3. An SFP effect did not occur for the Titled Discourse group. A centroparietal P600 effect emerged for Untitled Discourse group, but not until Critical Word 3. Rather, the SFP occurred earlier at Critical Word 2 for this group. In Study 1, participants' ability to identify the title of the discourses significantly increased after reading Critical Word 2 than Critical Word 1 ($p < .001$). In light of these findings, this may suggest that the SFP-effect occurring to Critical Word 2 represents a distinct post-lexical process of contextual ambiguity revision/resolution. Given that no centroparietal P600 effect was observed for the Untitled Discourse group until Critical Word 3, this may indicate that the P600 reflects discourse context updating only when a context is available. When a context is ambiguous, however, the context must first be resolved/revised (SFP effect at Critical Word 2) prior to being updated (centroparietal P600 effect at Critical Word 3).

These findings may explain the N400 effects. For example, SFP amplitudes were larger to Critical Word 2 than Critical Word 1 for the Untitled Discourse group, indexing possible ambiguity resolution at Critical Word 2. As such, the semantic aspects of Critical Word 3 were primed which was reflected as less negative N400 amplitudes to this word. N400 amplitudes were also larger to Critical Word 2 than Critical Word 1 for the Untitled Discourse group. This finding was unexpected given that lexical-semantic retrieval was hypothesized to not increase (smaller N400 amplitudes) until the context of the discourse was updated (increased P600 amplitudes).

The within-group P600- and SFP effects suggest that positive ERP components following the N400 reflect post-lexical processing during ambiguous discourse comprehension. However, the availability of a prior context influences the type and amount of post-lexical mechanisms that comprehenders may employ. That is, the ease of lexical sematic-retrieval (interpreted as less negative N400s) increased as the context of the discourses became apparent. This was represented by the P600 effect for Titled Discourse group, but rather an SFP effect for the Untitled Discourse group at Critical Word 2 and a P600 + SFP response for the Untitled Discourse group at Critical Word 3. Figure 4.6 summarizes the different patterns of N400, P600, and SFP effects for the Untitled Discourse and Titled Discourse groups throughout ongoing discourse comprehension.

Note the significant P600 effects for the Titled Discourse group at Critical Words 2 and 3, but for the Untitled Discourse group, only the SFP effect at Critical Word 2. The

*The woman went through her **list**.*

*She only had a limited amount of **room**.*

*She only had a limited amount of **room**.*

*She thought about getting a bigger **suitcase**.*

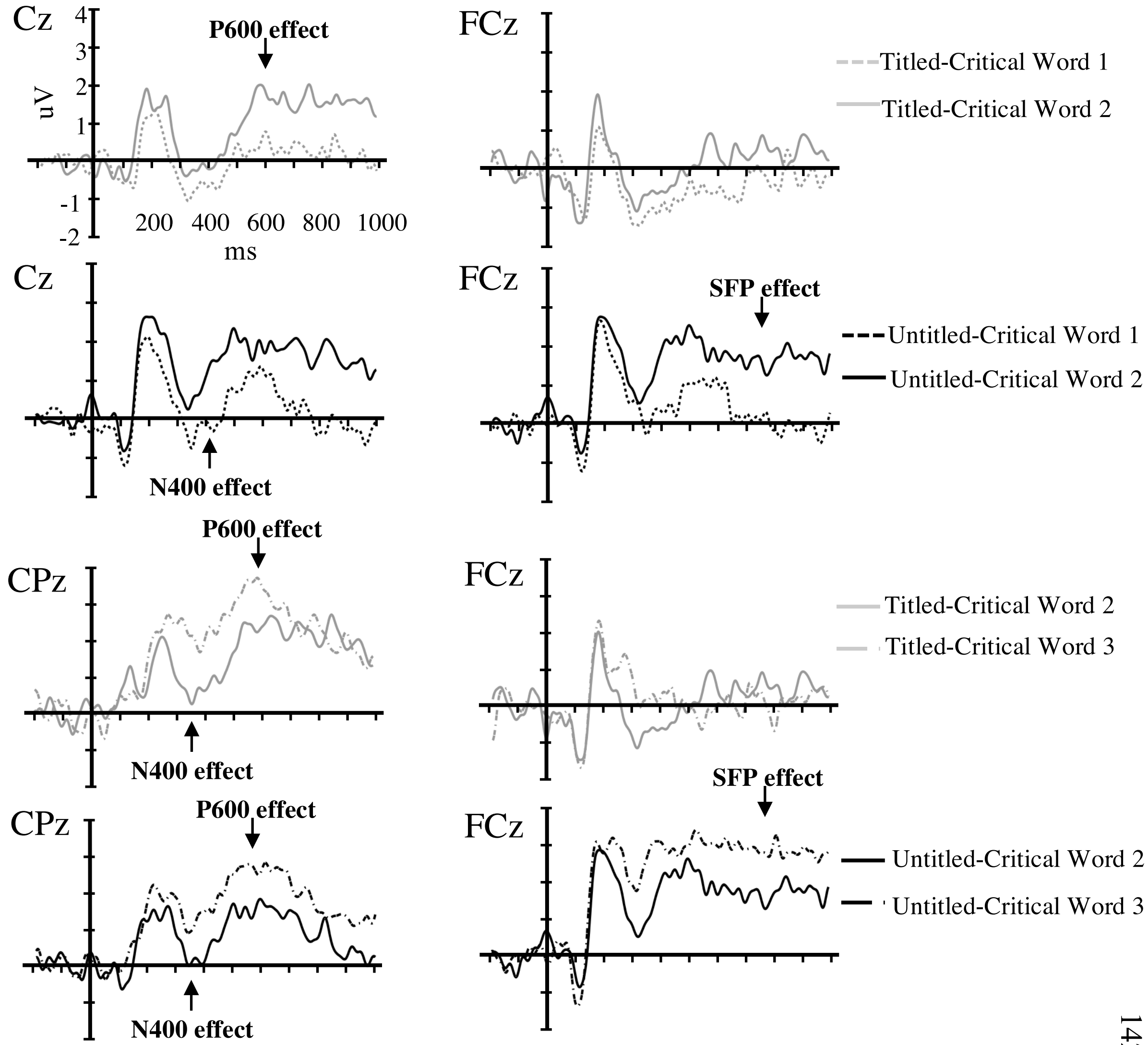

Figure 5.1. Patterns of N400, P600, and SFP effects at each Critical Word for each Group

arrows demark significant ERP effects. For example, the Titled Discourse group did not generate a difference in N400 amplitudes between Critical Words 1 and 2. However, this group generated a larger P600 to Critical Word 2 and a less negative N400 amplitude to Critical Word 3. This suggests that the Titled Discourse group, although aware of the contexts of the discourses, continued to engage in a post-lexical updating process to Critical Word 2 (increased P600). This post-lexical updating influenced the ease of lexical-semantic retrieval to Critical Word 3 (decreased N400) and further post-lexical updating to Critical Word 3 (increased P600).

The pattern of centroparietal P600 effects was different for the Untitled Discourse group. Specifically, there was no difference in centroparietal P600 amplitudes between Critical Words 1 and 2. However, a larger centroparietal P600 emerged to Critical Word 3 than Critical Word 2. Instead of a P600 at Critical Word 2, a larger SFP amplitude occurred. Specifically, SFP amplitudes at FCz significantly increased from Critical Word 1 to Critical Word 2 and Critical Word 3. Given that the Untitled Discourse group was not provided with a context prior to the discourses, the expected P600 peak may represent a fast-acting "checking" or "confirmation" process of context-updating, whereas the SFP may reflect the process of resolving/revising the contextual ambiguity of a discourse.

5.2. Functional significances of the N400, P600, and SFP Components

Current theories of language processing suggest that semantic processing involves two primary mechanisms that are structured either hierarchically or operate in parallel. Researchers supporting the parallel perspective (e.g., Hagoort, 2003, 2008) suggest that the N400 ERP reflects neural mechanisms supporting post-lexical processing (e.g., integration of word with contextual information). Other researchers theorize that the N400 represents an automatic lexical process (e.g., Kutas & Federmeier, 2000, 2011), whereas the P600 represents the post-lexical integration process (e.g., Brouwer & Hoeks, 2013; Brouwer et al., 2012).

Researchers reported that the presentation of new semantic elements (i.e., referents) within contexts elicited P600 effects at centroparietal (Burkhardt, 2006) and frontal electrode sites (Kaan et al., 2007). P600 effects were also recorded to inferences (Burkhardt, 2007), semantically anomalous verbs (Kim & Osterhout, 2005), and semantically congruent but contextually anomalous nouns (Nieuwland & Van Berkum, 2005). Recent theories of the functional significance of the P600 suggest that it reflects the updating of the context or mental representation of a discourse (Brouwer et al., 2012). However, these reported P600 effects occurred in discourses in which a preceding context was always available and known. In contrast, the SFP reported in the current study occurred to words (Critical Word 2, Critical 3) in which the contexts were initially unknown and ambiguous. Study 1 established that participants' identification of the contexts of the 25 ambiguous discourses increased after reading Critical Word 2 and again after reading Critical Word 3. In light of these findings, it is proposed that the SFP reported in the current study reflects a post-lexical process distinct from domain-general context-updating (i.e., P600; Coulson, King, & Kutas, 1998), but instead reflects the resolution/revision of the contextual ambiguity in coherent discourse.

The current study challenges existing theories of language processing. Instead it proposes that discourse comprehension involves three distinct processes: (a) semantic information of the word is retrieved from the mental lexicon (N400; Kutas & Federmeier, 2000), (b) the new semantic information is quickly "checked" against the existing context (if a context is available) and updated (P600; Brouwer et al., 2012), and (c) the context of a discourse, if ambiguous, is partially resolved (Late Sustained Frontal Positivity). The second goal of this study was to empirically examine this theory and explore the functional significances of these ERP components.

As previously reviewed at length, researchers theorize that the N400 represents an automatic process of lexical-semantic retrieval (Federmeier & Kutas, 1999; Kutas & Federmeier, 2000) and the P600 represents a process in which the mental model of a discourse is updated (e.g., Brouwer et al., 2012). A more likely explanation is that there is a "family" of P600 effects that may reflect some shared, underlying neural processing, including syntactic revision (Friederici, Hahne, & Saddy, 2002; Hagoort, 2003, 2008), discourse context-updating (Brouwer et al., 2012; Burkhardt, 2007), and conflict monitoring (Van de Meerendonk et al., 2001). Others suggest that the P600 reflects a domain-general process that is generated to unexpected events (Coulson et al., 1998).

The current findings support the theory that the P600 reflects a post-lexical process occurring during discourse comprehension. It is presumed that the ability to update the context of a discourse is influenced by working memory of the previously presented contextual information. Given these theoretical perspectives of the N400 and P600 ERPs, the current study hypothesized (Hypothesis 3) that a behavioral measure involving lexical-semantic retrieval (The Boston Naming Task) moderated N400 amplitudes. Similarly, it was hypothesized (Hypothesis 4) that P600 amplitudes would be influenced by an established measure of working memory during reading (Reading Span Task). Given the unexpected SFP effects in the Untitled Discourse group to Critical Word

2 and Critical Word 3, I further explored if performance on the Reading Span Task influenced SFP amplitudes.

Hypothesis 3 was partially supported. Pairwise comparisons indicated that the moderating effect of BNT accuracy and response time was different for the Untitled Discourse and Titled Discourse groups. Specifically, when BNT accuracy was at the mean and +1 SD above the mean, the Untitled Discourse group generated significantly less negative N400s to Critical Word 3 than Critical Words 1 and 2 at both CPz and Cz electrode clusters. However, when BNT accuracy was -1 SD below the mean, this relationship was only visible at Cz clusters for the Untitled Discourse group. The same pattern of results emerged when BNT response time was at slow, medium, and fast- levels: When response time was slow, the Untitled Discourse group only generated N400 amplitude differences between words at Cz, but not CPz. Based on these findings, I conclude that while reading a story in the absence of a context, participants with worse confrontation naming abilities may engage fewer neural resources to support lexical-semantic retrieval.

In partial support of Hypothesis 4, SFP, but not P600, amplitudes varied based on the availability of a prior context (Untitled Discourse, Titled Discourse) and working memory capacity. To review, irrespective of Reading Span, centroparietal P600 amplitudes did not vary between Critical Words 1 and 2. However, across all three levels of Reading Span, centroparietal P600 amplitudes were larger to Critical Word 3 than Critical Word 2. These effects extended into the SFP-latency window when Reading Span was low or at the mean, but not when Reading Span was above the mean. Regardless of Reading Span, larger SFP amplitudes occurred to Critical Word 2 than

Critical 1. When Reading Span was average or above average, but not low, these frontocentral "early latency" SFP effects began to emerge during the P600 latency window. Therefore, individuals with low working memory capacity generated a later-occurring SFP effect than those with average or above average working memory. These findings may suggest that working memory capacity moderates the onset of the SFP. In review, at Critical Word 2, the frontocentral SFP effect emerged in both the P600 and SFP latency windows when Reading Span was average or above average. However, when Reading Span was low, the SFP effect at Critical Word 2 did not emerge until the later SFP latency window. Also, the P600 effect to Critical Word 3 extended into the SFP latency window when working memory capacity was average or below, but not high. This suggests that when working memory capacity is high, individuals may employ a quicker ambiguity resolution process.

Surprisingly, amplitudes in the SFP time window occurred to Critical Word 3 and Critical Word 2 than Critical Word 1 at midline Cz for the Titled group. However, follow-up analysis suggested that these results only occurred when Reading Span was at low and average levels. It is likely that the larger SFP amplitudes at Cz for the Titled Discourse group reflect less peak-like P600 effects, as ERPs dissipate laterally from their location of maximum scalp recording. On the other hand, this effect may suggest that even when a context was provided for some participants (Titled Discourse group), a low-level of contextual ambiguity may have remained when reading the discourses. This may be an indirect effect of the filler discourses that included semantic anomalies. Because of the semantic anomalies, participants in the Titled Discourse group may have learned throughout testing that some discourses would not follow the contexts provided for them.

As a result, these participants may have learned to become skeptical of the “true” context of a discourse. However, this effect only occurred at Cz for the Titled group, but for the Untitled group, was maximal at frontal and frontocentral electrode cluster. Thus, it is unlikely that this effect at Cz for the Titled Discourse group reflects psycholinguistic processing underlying the SFP effect for the Untitled Discourse group. This is because no SFP effect was observed between Critical Words at the frontal/frontocentral electrode clusters—the sites where the SFP was maximal.

This pattern of P600- and SFP effects when Reading Span was constant at different levels further supports the theory that the P600 represents an early, efficient post-lexical updating process, whereas the SFP reflects a separate ambiguity resolution/revision process. For clarity, the results that support this theory are summarized.

First, the Titled Discourse group generated P600 effects to Critical Word 2 and Critical Word 3. This suggests that even when contextual information was provided (i.e., a title), individuals engage in a post-lexical process as represented by the P600. It is possible that, when a context is available and known, all words generate a N400-P600 complex (Brouwer et al., 2012).

Second, the Untitled Discourse group only generated a centroparietal P600 at Critical Word 3. However, larger frontocentral SFP amplitudes emerged both at Critical Word 2 and Critical Word 3. In contrast, SFP effects did not occur for the Titled Discourse group at frontal/frontocentral electrode channels for any Critical Word. Taken together, this suggests that, overall, the Untitled Discourse group engaged in a different additional type of post-lexical cognitive process at Critical Words 2 and 3. Given that larger P600 amplitudes occurred for this group at Critical Word 3 only, a context may have to be available (reflected by larger SFP at Critical Word 2) in order to engage in context updating processing (P600 at Critical Word 3).

Third, working memory capacity moderated the timing of SFP effects for the Untitled Discourse group. When working memory was average or above average, but not low, frontocentral SFP-effects emerged earlier in the P600 latency window in addition to the later SFP latency window. This suggests that individuals with low working memory capacity may employ delayed ambiguity resolution processing early in an ambiguous discourse. However, working memory capacity did not moderate the centroparietal P600.

The current evidence suggests that when reading coherent discourse, individuals employ post-lexical processing whether or not an available context was provided. When a context was provided (Titled Discourse group), P600 effects emerged as the discourse unfolded for those with average to above average working memory capacity. When a context was not provided (Untitled Discourse group), individuals first utilized cognitive processes, represented by the SFP, that my represent contextual ambiguity resolution. The cognitive process that the P600 represents (e.g., discourse context-updating) may only emerge after contextual ambiguity is resolved or revised (SFP effect at Critical Word 2; P600 and SFP effect at Critical Word 3).

To the author's knowledge, this study is the first to report the presence of a sustained frontal positive ERP component recorded to words in contextually ambiguous discourses. Other research reported on a Late Positive Component (LPC) that was suggested to reflect increased context updating to the establishment of a new discourse referent (Kaan et al., 2007). Similar to the SFP reported in the current study, the LPC was distributed across scalp recording sites. However, the onset of the LPC was much later (approximately 900 ms) than the SFP reported in the currents study.

Kolk and colleagues (2003) reported on a slow positive shift with similar latency to the SFP (approximately 600-1000 ms post-stimulus onset) for semantically and syntactically unacceptable verbs in contextually unambiguous sentences. The amplitude of the slow positive shift was maximal at central and parietal electrodes, in contrast to the frontocentral maximum of the SFP in the current study. Other research, however, reported on a frontal P600 (FP600) that was larger to verbs in ambiguous two noun-phrase sentences than unambiguous one noun-phrase sentences (Kaan & Swaab, 2003). This positivity was largest at right frontal electrode recording sites between 500 and 900 ms post critical-verb onset. Those researchers concluded that this frontal positivity reflected an increase in discourse complexity due to the more ambiguous two noun-phrase sentences, whereas the posterior P600 reflects a simpler integration process. Although this theory resembles that which is currently proposed in this study on the functional significances of the P600 and SFP components, the topography of the FP600 differed from the SFP. Specifically, the FP600 was characterized by a clear “peak” similar to that of the posterior P600. In contrast the SFP does not contain a peak, but rather is a frontal positivity that is sustained through 1000 ms post stimulus onset and possibly even further (see Figures 4.2 and 4.5). Therefore, it is possible that the SFP reflects combinatory frontal mechanisms (e.g., establishment of a new discourse referent, revision, ambiguity resolution) rather than a single isolated process.

Others also dissociated centroparietal P600 effects from that which is frontally distributed. Specifically, sentences contained words that either required syntactic repair or more complex syntactic processing (Friederici et al., 2002). The researchers found that syntactically incorrect words elicited a centroparietal P600 effect, whereas words of greater syntactic complexity elicited a frontocentral P600 effect. This study concluded that different neural mechanisms underlie syntactic repair and syntactic complexity. However, this study was limited to examining syntactic, but not semantic, P600 effects. In summary, the polarity, morphology, and scalp distribution of the SFP suggests that it reflects a cognitive process or group of cognitive processes that are different than those which reflect the P600. In light of the experimental manipulations established in Study 1, the SFP is interpreted as reflecting the resolution of contextual ambiguity within coherent discourses.

5.3. Limitations and alternative explanations

This study is not without limitations. First, one must acknowledge possible varying levels of arousal between the Titled Discourse and Untitled Discourse groups during ERP recording (Luck, 2014). Although the stimuli were identical for the two groups, task demands were undoubtedly greater for participants in the Untitled Discourse group who needed to actively search for the contexts of the discourses. As a result, it is possible that the SFP effects in the Untitled Discourse group may be due to higher levels of arousal than the Titled Discourse group at Critical Word 1 and Critical Word 2. The SFP ERP may also reflect some domain-general learning process that may not be specific to discourse comprehension. For example, given the high levels of task demands for the Untitled Discourse group (i.e., actively search for the meaning of a discourse), the SFP may reflect increased arousal due to the excitement associated with ambiguity resolution and/or problem solving.

Second, concreteness (Brysbaert et al., 2014) of Critical Word 3 was greater than Critical Word 1 in the ERP task. Researchers suggested the N400 amplitudes were larger to concrete words than abstract words at frontal electrode channels (e.g., Barber et al., 2013). However, all Critical Words were at least “moderately concrete” (Brysbaert et al., 2014). In addition, any effect of concreteness on the centroparietal N400 was equal across Untitled Discourse and Titled Discourse groups because both groups were exposed to exactly the same stimuli. Thus, the larger concreteness to Critical Word 3 than Critical Word 1 does not influence the interpretation of the current study’s findings.

Third, it is important to recognize that the current study reported findings from group-averaged ERPs. Thus, findings may not apply to all individuals. Although this study attempted to identify individual differences that influence language processing (i.e., confrontation naming abilities; working memory capacity), there are likely other mechanisms of individual variability. Established behavioral measures of confrontation naming (Boston Naming Test) and working memory capacity (Reading Span Task) were used, but these also require other perceptual (e.g., object recognition) and cognitive processing (e.g., attention) skills which may have influenced the ERPs examined in the current study.

The current findings were dependent on participants’ interpretation of context and the perception of potential sources of ambiguity within a context even when a context was provided (i.e., Titled Discourse group). Thus, it is necessary to examine how predisposing factors (e.g., biological sex, gender, race, ethnicity) influence the interpretation of context during discourse comprehension and associated underlying neural mechanisms.

The data collection sequence of Study 1 and Study 2 was critical to the results of the overall program of this research. Specifically, Study 1 established that participants did not identify the topics of the 25 discourses after sentence 1, but that title identification increased after reading sentence 2 and even more so after sentence 3. This result provided behavioral evidence that guided the interpretation of findings from Study 2. Specifically, that P600 and SFP effects in the Untitled Discourse group represented two different sub-mechanisms of post-lexical processing. However, it is possible that differences in automatic and controlled lexical processing occurred to other words in the discourses than the last words in each sentence. If so, the current findings of the Critical Words examined in the present study may be underestimates of effects that were maximal at earlier points in the sentences.

The theory proposed here assumes that contextual information must be available in order for increased lexical-semantic retrieval abilities to occur (St. George et al., 1994). This would propose that, when no contextual information was available (Untitled Discourse group) N400 amplitudes would only become more positive after context-updating processing (P600 and/or SFP). However, the Untitled Discourse group generated more negative N400 amplitudes to Critical Word 2 than Critical Word 1.

There are three possible explanations for this effect. First, increased ease of lexical-semantic retrieval may occur independent of post-lexical processing. Said otherwise, spreading activation to semantically-related nodes in semantic memory (Collins & Loftus, 1975) may occur irrespective of whether a context is available or not.

Second the N400 may reflect, at least in part, aspects of both automatic and controlled lexical processing (Holcomb et al., 2005; Rolke et al., 2001). This may explain why the Untitled Discourse group generated a less negative N400 and more positive SFP to Critical Word 2 than Critical Word 1. Third, it is possible that the Untitled Discourse group engaged in contextual ambiguity revision (represented by a larger SFP amplitude) earlier in the discourses than Critical Word 2. This may explain why an unexpected less negative N400 amplitude to Critical Word 2 than Critical Word 1 occurred for the Untitled Discourse group.

Finally, the temporal window of ERP analysis (through 1000 ms) limited the interpretation of the time-course of the Late Sustained Frontal Positivity. Visual inspection of Figure 4.2 suggests that the latency of the SFP may extend beyond 1000 ms.

5.4. Conclusions and future directions

Two studies were performed to examine the organization of the language processor and investigate the functional significances of the N400 and P600 event-related potentials during the comprehension of contextually ambiguous coherent discourses. The present findings support theories suggesting that the N400 represents an automatic process, such as lexical-semantic retrieval. On the other hand, the amplitude of the P600 reflected, at least in part, post-lexical processing, such as discourse context updating. Unexpectedly, for the Untitled Discourse group, a Late Sustained Frontal Positivity (SFP), but not P600, occurred when the contexts of the discourses started to become more clear (Critical Word 2). To Critical Word 3, however, P600- and SFP effects emerged for this group. Study 1 found that participants' ability to identify the title of the contextually ambiguous discourses increased after reading Critical Words 2 and 3 than Critical Word 1 (ps < .001). Considering these findings in conjunction with those from Study 2, it is suggested that the SFP reflects the resolution/revision of contextual ambiguity, whereas the P600 reflects discourse context updating occur when an existing context is available.

Working memory capacity moderated the timing of the SFP effect. Specifically, when working memory capacity was average or above, the SFP effect emerged earlier (P600 latency window) than when working memory capacity was low (SFP latency window only). A novel theory of the organization of the language processor during discourse comprehension was proposed to account for the present studies' results. This theory suggests that lexical information of the word is first retrieved from semantic memory (N400). Following, if a discourse context is available, a fast-acting context-updating process occurs (P600). If a context is ambiguous, however, it must first be revised/resolved (SFP) before discourse context updating may occur.

The present findings established a temporal sequence of language processing that occurs during naturalistic discourse comprehension. A temporal framework, such as the one proposed, will provide neurologists and neuropsychologists with more targeted approaches for examining the neural networks that support discourse comprehension which may be compromised following brain injury. Additional research using more spatially sensitive brain imaging techniques (e.g., MEG, fMRI, ERP source localization) is necessary to identify the neural origins of the N400, P600, and SFP ERPs and associated neural pathways (e.g., diffusion tensor imaging).

These findings are also critical for improving our understanding of the different neural mechanisms that may be compromised in individuals with cognitive-communication disorders (e.g., traumatic brain injury, dementia, right hemisphere dysfunction). Further research is needed to understand how cognitive-communication interventions may differentially influence the amplitudes of the N400, P600, and SFP ERP components. These studies are important for developing more targeted interventions to improve functional cognitive and communication outcomes in clinical populations.

REFERENCES

Aborn, M., Rubenstein, H., & Sterling, T. D. (1959). Sources of contextual constraint upon words in sentences. Journal of Experimental Psychology, 57(30), 171-180.

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: SAGE Publications.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39(3), 445-459.

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. Brain & Language, 125, 47-53.

Barrett, S. E., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. Neuropsychologia, 27(7), 913-922.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. Brain and Cognition, 14, 201-212.

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event-related potentials and the matching of familiar and unfamiliar faces. Neuropsychologia, 26(1), 105-117.

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. Journal of Experimental Psychology: Human Perception and Performance, 21(1), 54-67.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. Electroencephalography and Clinical Neurophysiology, 60(4), 343-355.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. Memory & Cognition, 8(6), 631-642.

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. Behavior Research Methods, 42(3), 665-670.

Bransford, J. D., & Johnson, M. C. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

Brouwer, H., Fitz, H., & Hoeks, J. (2012).Getting real about Semantic Illusions: Rethinking the role of the P600 in language comprehension. Brain Research, 1446, 127-143.

Brouwer, H., & Hoeks, J. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. Frontiers in Human Neuroscience, 7, 1-12.

Brown, B., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. Journal of Cognitive Neuroscience, 5(1), 34-44.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41(4), 997-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904-911.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. Brain and Language, 98, 159-168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. NeuroReport, 18(17), 1851-1854.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience, 12(1), 1-47.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. Psychophysiology, 32, 274-285.

Clark, H. H., & Haviland, S. E. (1974). Psychological processes as linguistic explanation. In D. Cohen (Ed.), Explaining Linguistic Phenomena (pp. 91-124). Washington, D.C.: Hemisphere Publishing Corp.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. Psychological Review, 82(6), 407-428.

Cortese, M. J., & Balota, D. A. (2012). Visual word recognition in skilled adult readers. In M. Spivey, M. Joanisse, & K. McRae (Eds.), The Cambridge Handbook of Psycholinguistics (pp. 159-185). New York, NY: Cambridge University Press.

Costafreda, S. G., Fu, C. H. Y., Lee, L., Everitt, B., Brammer, M. J., & David, A. S. (2006). A systematic review and quantitative appraisal of fMRI students of verbal fluency: Role of the left inferior frontal gyrus. Human Brain Mapping, 27, 799-810.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. Language and Cognitive Processes, 13(1), 21-58.

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. Psychophysiology, 41, 60-74.

Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not represent a post-lexical process. Cognitive Brain Research, 9, 137-146.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience, 8(8), 1117-1121.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: a review and tutorial, In T. Handy (Ed.), Event Related Potentials: A Methods Handbook (pp. 57-82). Cambridge, MA: MIT Press.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. Journal of Experimental Psychology, 88(2), 216-222.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. Quarterly Journal of Experimental Psychology, 33A, 325-350.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. Journal of Memory and Language, 41, 469-495.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. The Psychology of Learning and Motivation, 51, 1-44.

Federmeier, K. D., Wlotka, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. Brain Research, 1146, 75-84.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potential related to stages of sentence verification. Psychophysiology, 20(4), 400-409.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts of naming time: Evidence for autonomous lexical processing. Quarterly Journal of Experimental Psychology, 33A, 465-495.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 12, 627-635.

Francis, W. N. & Kučera, H. (1982). Frequency Analysis of English Usage: Lexicon and Grammar. Boston, MA: Houghton Mifflin Company.

Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. NeuroReport, 15(1), 165-169.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. Journal of Psycholinguistic Research, 31(1), 45-63.

Grossi, G. (2006). Relatedness proportion effects on masked associative priming: An ERP study. Psychophysiology, 43, 21-30.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. Journal of Cognitive Neuroscience, 15(6), 883-899.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. Philosophical Transactions of the Royal Society of Britain, 363, 1055-1069.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes, 8(4), 439-483.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. Clinical Neurophysiology, 115, 1090-1103.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. Cognitive Brain Research, 19, 59-73.
Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(3), 721-742.
Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. Language and Cognitive Processes, 5(4), 281-312.
Holcomb, P. J., Reder, L., Misra, M., & Grainger, J. (2005). The effects of prime visibility on ERP measures of masked priming. Cognitive Brain Research, 24(1), 155-172.
Junghöfer, M., Elbert, T., Leiderer, P., Berg, P., & Rockstroh, B. (1997). Mapping EEG-potentials on the surface of the brain: A strategy for uncovering cortical sources. Brain Topography, 9, 203-217.
Kaan, E., Dallas, A. C., Barkley, C. M. (2007). Processing bare quantifiers in discourse. Brain Research, 1146, 199-209.
Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. Journal of Cognitive Neuroscience, 15(1), 98-110.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General, 133(2), 189-217.
Kaplan, E., Goodglass, H., & Weintraub, S. (2001) Boston Naming Test. Pro-ed.
Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., . . . Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology, 51, 1-21.
Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. Cognitive Brain Research, 13, 27-39.
Kiefer, M., & Spitzer, M. (2000). Time course of conscious and unconscious semantic brain activation. NeuroReport, 11, 2401-2407.
Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. Journal of Memory and Language, 52, 205-225.
Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. Discourse Processes, 39(2&3), 125-128.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95(2), 163-182.
Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension, Psychologia, 42, 186-198.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85(5), 363-394.
Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. Brain and Language, 85, 1-36.
Kotz, S. A., Frisch, S., Von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. Journal of the International Neuropsychological Society, 9, 1053-1060.
Kounious, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(4), 804-823.
Kučera, H. & Francis, W. N. (1967). Computational Analysis of Present-Day American English. Providence, RI: Brown University Press.
Kutas, M., DeLong, K., & Kiang, M. (2011). From phonemes to discourse: Event-related brain potential (ERP) componentry and paradigms for investigating normal and abnormal language processing. In J. Guendouzi, F. Loncke, & M. Williams (Eds.), The Handbook of Psycholinguistic and Cognitive Processes (pp. 119-144). New York, NY: Psychology Press.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences, 4(12), 463-470.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential. Annual Review of Psychology, 62, 621-647.
Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. Journal of Cognitive Neuroscience, 1(1), 38-49.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Letters to Nature, 307(12), 161-163.
Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. Memory & Cognition, 11(5), 539-550.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205.
Kutas, M. & Van Petten, C. (1994). Psycholinguistics Electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics (pp. 83-143). San Diego, CA: Academic Press, Inc.
Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. Electroencephalography and Clinical Neurophysiology, 69, 218-233.
Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. Journal of Memory and Language, 61(3), 326-338.
Lau, E. F., Almeida, D., Hines, P., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. Brain and Language, 111(3), 161-172.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. Nature Reviews Neuroscience, 9, 920-933.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. TRENDS in Cognitive Sciences, 10(10), 447-454.

Luck S. J. (2014). An introduction to the event-related potential technique (2nd ed). Cambridge, MA: The MIT Press.

Luck, S. J., Vogel, E., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. Letters to Nature, 383, 616-618.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. Psychological Methods, 7(1), 19-40.

Manly, B. F. (2004). Multivariate statistical methods: A primer. (3rd ed.). Boca Raton, FL: CRC Press.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. Psychological Bulletin, 113(1), 181-190.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities of auditory event-related potentials. Electroencephalography and Clinical Neurophysiology, 59, 477-488.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(4), 919-932.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. Psychophysiology, 36, 53-65.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90(2) 227-234.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. Discourse Processes, 26(2&3), 131-157

Näätänen, R. (1970). Evoked potential, EEG, and slow potential correlates of selective attention. Acta Psychologia, 33, 178-192.

Näätänen, R. (1967). Selective attention and evoked potentials. Thesis, University of Helsinki, Finland.

Nestor, P. G., Kimble, M. O., O'Donnell, B. F., Smith, L., Niznikiewicz, M., Shenton, M. E., & McCarley, R. W. (1997). Aberrant semantic activation in schizophrenia: A neurophysiological study. American Journal of Psychiatry, 154(5), 640-646.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractioning language: Different neural subsystems with different sensitive periods. Cerebral Cortex, 2, 244-258.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Cognitive Brain Research, 24, 691-701.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience, 18(7), 1098-1111.

Nigam, A., Hoffman, J. E., & Simon, R. F. (1992). N400 to semantically anomalous pictures and words. Journal of Cognitive Neuroscience, 4(1), 15-22.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. Brain and Cognition, 62, 267-272.
Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. Neuroscience Letters, 436(2), 133-137.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Memory and Language, 31, 785-806.
Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. BMC Neuroscience, 8(1), 1-9.
Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. Brain Research, 1153, 166-177.
Pinheiro, A. P., Del Re, E., Nestor, P. G., Mezin, J., Rezaii, N., McCarley, R. W., . . . Niznikiewicz, M. (2015). Abnormal interactions between context, memory structure, and mood in schizophrenia: An ERP investigation. Psychophysiology, 52, 20-31.
Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage, 10, 15-35.
Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium (pp. 669–682). Hillsdale, NJ: Lawrence Erlbaum.
Pratarelli, M. E. (1995). Modulation of semantic processing using word length and complexity: an ERP study. International Journal of Psychophysiology, 19, 233-246.
Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological Review, 95(3), 385-408.
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? Journal of Experimental Psychology: Human Perception and Performance, 18(3), 849-860.
Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. European Journal of Psychological Assessment, 28(3), 164-171.
Regan, D. (1972). Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman and Hall Ltd.
Revonsuo, A., Portin, R., Juottonen, K., & Rinne, J. O. (1998). Semantic processing of spoken words in Alzheimer’s Disease: An electrophysiological study. Journal of Cognitive Neuroscience, 10(3), 408-420.
Rolke, B., Heil, M., Streb, J., & Henninghausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. Psychophysiology, 38, 165-174.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. Memory & Cognition, 18(4), 367-379.
Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. Psychophysiology, 22(6), 642-647.
Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. Journal of Cognitive Neuroscience, 23(3), 514-523.
Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. Proceedings of the National Academy of Science, 95, 12061-12068.
Song, J., Davey, C., Poulsen, C., Turovets, S., Luu, P., & Tucker, D. M. (2014). Sensor density and head surface coverage in EEG source localization, In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI 2014) (pps. 620–623). IEEE: New York.
St. George, M., Kutas, M., Martinez, A., & Sereno, A. I. (1999). Semantic integration in reading: engagement of the right hemisphere during discourse processing. Brain, 122(7), 1317-1325.
St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. Journal of Cognitive Neuroscience, 6(1), 70-83.
St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. Journal of Cognitive Neuroscience, 9(6), 776-787.
Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. Science, 150(3700), 1187-1188.
Swick, D., Kutas, M., & Knight, R. T. (1998). Prefrontal lesions eliminate the LPC but do not affect the N400 during sentence reading. Journal of Cognitive Neuroscience Supplement, 5, 29.
Taylor, A. (2011). Using the GLM Procedure in SPSS. Student Handbook. Macquarie University.
Taylor, W. L. (1953). “Cloze Procedure”: A new tool for measuring readability. Journalism Quarterly, 30, 415-433.
Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent variable analysis of the relationship between processing and storage. Memory, 17(6), 635-654.
Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. Clifton (Eds.), The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond (pp. 229-270). New York, NY: Psychology Press.
Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.
Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. Journal of Cognitive Neuroscience, 11(6), 657-671.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to a wider discourse? Evidence from the N400 effect. Cognitive Brain Research, 17, 701-718.

Van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. NeuroImage, 54, 2350-2363.

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

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related potentials. Memory & Cognition, 18(4), 380-393.

Van Petten, C. & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language, 97, 279-293.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related potential measures. Neuropsychologia, 33(4), 485-508.

Van Petten, C., Weckerly, J., McIssac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. Psychological Science, 8(3), 238-242.

Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. Language and Cognitive Processes, 16(1), 65-103.

Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., & Gabrieli, J. D. E. (1997). Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. Journal of Cognitive Neuroscience, 9(6), 714-726.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. Cognitive Brain Research, 13, 363-375.

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. Journal of Cognitive Neuroscience, 20(7), 1235-1249.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology, 42, 1-15.

Xiao, X., Zhao, D., Zhang, Q., & Guo, C. (2012). Retrieval of concrete words involves more contextual information than abstract words: Multiple components for the concreteness effect. Brain & Language, 120, 251-258.

Yurchenko, A., den Ouden, D. B., Hoeksema, J., Dragoy, O., Hoeks, J. C. J., & Stowe, L. A. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. Neuropsychologia, 51, 132-141.

APPENDICES

APPENDIX A: DESCRIPTIVE STATISTICS OF NARRATIVE DISCOURSES

	Title	Sent.[1]	Words[2]	CW[3]	Lett.[4]	Syll.[5]	Conc.[6]	Freq.[7]	Cloze (U)[8]	Cloze (T)[9]
1	Doing laundry	1	7	piles	5	1	4.19	91	10.3	6.9
		2	7	buttons	7	2	4.80	444	0.0	0.0
		3	8	washer	6	2	4.7	104	3.3	81.8
2	Taking a driving test	1	10	training	8	2	3.30	1792	0.0	3.4
		2	7	instructor	10	3	4.26	240	0.0	67.86
		3	6	car	3	1	4.89	24636	0.0	54.5
3	Making a peanut butter and jelly sandwich	1	9	mess	4	1	3.90	3985	0.0	3.4
		2	6	jars	4	1	5.0	99	0.0	0.0
		3	8	bread	5	1	4.92	1445	3.3	78.8
4	Going to the movies	1	6	counter	7	2	4.17	709	0.0	3.4
		2	8	cash	6	2	4.48	3694	3.2	0.0
		3	7	popcorn	7	2	5.0	465	0.0	87.9
5	Wedding ceremony	1	9	flowers	7	1	5.0	2427	0.0	10.3

		2	7	attendants	10	3	4.44	56	0.0	0.0
		3	6	bride	5	1	4.63	1235	10.0	84.8
6	Eating ice cream	1	8	outdoors	8	2	4.61	107	0.0	3.4
		2	9	bowl	4	1	4.87	1094	0.0	14.3
		3	7	vanilla	7	3	4.68	203	0.0	54.5
7	Kids getting on the bus	1	6	beep	4	1	4.0	332	0.0	0.0
		2	7	house	5	1	5.0	26214	3.2	42.9
		3	8	bus	3	1	4.90	3783	3.3	57.6
8	Stuck outside in the rain	1	6	head	4	1	4.75	18947	0.0	69.0
		2	9	television	10	4	4.83	1729	0.0	0.0
		3	7	weather	7	2	3.83	1746	10.0	54.5
9	Gambling	1	7	prize	5	1	4.45	1142	10.3	13.8
		2	8	sunglasses	10	3	4.83	201	0.0	3.6
		3	8	cards	5	1	4.90	2449	3.3	54.5
10	Learning to ride a bike	1	6	repetitions	11	4	3.17	5	0.0	0.0
		2	7	scratches	9	2	4.59	123	0.0	10.7
		3	7	pedals	6	2	4.44	38	3.3	90.91
11	Packing for a trip	1	6	list	4	1	4.43	4110	0.0	3.4
		2	8	room	4	1	4.79	22415	0.0	7.1

		3	7	suitcase	8	2	4.97	683	6.7	78.8
12	Brewing coffee	1	7	kitchen	7	2	4.97	2974	0.0	44.8
		2	7	droplet	7	2	4.45	2	0.0	39.3
		3	7	mug	3	1	4.80	349	10.0	63.6
13	Decorating the Christmas tree	1	8	doorway	7	2	4.75	164	0.0	0.0
		2	8	ladder	6	2	5.0	472	0.0	7.1
		3	8	star	4	1	4.69	4149	13.3	48.5
14	Eating an orange	1	10	fingernails	11	3	4.93	210	0.0	3.4
		2	6	smell	5	1	3.70	4240	0.0	14.3
		3	8	peel	4	1	4.61	273	10.0	66.7
15	Playing the guitar	1	6	grip	4	1	4.30	494	0.0	0.0
		2	6	audience	8	2	4.29	1294	0.0	0.0
		3	6	strings	7	1	4.75	418	0.0	51.51
16	Having a picnic	1	9	months	6	1	4.20	8314	3.4	6.9
		2	9	clearing	8	2	3.79	331	0.0	0.0
		3	7	blanket	7	2	5.0	662	13.3	81.8
17	Graduation	1	8	outfit	6	2	4.12	1280	0.0	13.8
		2	8	bleachers	9	2	4.14	59	12.9	21.4

		3	7	diploma	7	3	4.93	129	3.3	96.7
18	Mouse in the house	1	6	hunt	4	1	3.81	1319	0.0	3.4
		2	7	intruder	8	3	4.14	186	0.0	0.0
		3	8	cheese	6	1	4.70	1991	6.7	81.8
19	At a NASCAR race	1	8	fans	4	1	4.71	711	3.4	51.7
		2	7	ears	4	1	5.0	1705	12.9	42.9
		3	6	cars	7	2	4.71	340	3.3	51.5
20	Listening to the car radio	1	6	ride	4	1	3.75	6904	0.0	0.0
		2	7	switch	6	1	4.07	1434	0.0	0.0
		3	7	station	7	2	4.32	4033	10.0	84.8
21	Getting a haircut	1	6	calendar	8	3	4.62	363	0.0	0.0
		2	7	party	5	2	3.89	11890	9.7	14.3
		3	6	trim	4	1	3.93	218	0.0	45.5
22	Getting in a fight	1	7	arms	4	1	4.97	3050	0.0	0.0
		2	6	soreness	8	2	3.50	5	0.0	0.0
		3	7	punch	5	1	4.39	1514	0.0	60.6
23	Gardening	1	5	holes	5	1	4.81	779	0.0	41.4
		2	6	jeans	5	1	5.0	337	0.0	0.0
		3	6	seeds	5	1	4.71	228	3.3	45.45

24	Opening the front door	1	7	mat	3	1	4.83	178	0.0	6.9
		2	6	bag	3	1	4.90	4796	19.4	3.6
		3	6	key	3	1	4.89	4430	10.0	78.79
25	Teacher writing on a chalkboard	1	6	wall	4	1	4.86	3605	0.0	0.0
		2	7	marks	5	1	4.21	1117	0.0	0.0
		3	6	students	8	2	4.92	1564	10.0	63.64

Note. [1]Sentence within discourse; [2]Words per sentence; [3]Critical Word; [4]Number of letters of Critical Word; [5]Number of syllables of Critical Word; [6]Conceretness of Critical Word (Brysbaert et al., 2014), [7]Frequency of Critical Word in English Lexicon (Brysbaert & New, 2009), [8]Cloze probability of Critical Word in Untitled conditions, [9]Cloze probability of Critical Word in Titled conditions. These discourses were retained based on the following cloze probability inclusion criteria: (a) < 20% cloze-probability of all critical words in Untitled condition, (b) > 45% cloze-probability of critical word 3 in Titled condition.

APPENDIX B

DESCRIPTIVE STATISTICS FOR TITLE IDENTIFICATION ACCURACY (%) FOR EACH WAVE

	Discourse Title	Wave	Accuracy
1	Doing laundry	1	3.4
		2	0.0
		3	6.7
		4	79.3
2	Taking a driving test	1	0.0
		2	0.0
		3	0.0
		4	86.2
3	Making a peanut butter and jelly sandwich	1	0
		2	0
		3	3.3
		4	72.4
4	Going to the movies	1	0.0
		2	0.0
		3	0.0
		4	79.3
5	Wedding ceremony	1	3.4
		2	3.2
		3	6.7
		4	86.2
6	Eating ice cream	1	0.0
		2	0.0
		3	0.0
		4	86.2
7	Kids getting on the bus	1	3.4
		2	0.0
		3	6.7
		4	86.2
8	Stuck outside in the rain	1	0.0
		2	6.5
		3	10.0
		4	75.9
9	Gambling	1	0.0
		2	0.0
		3	3.3

		4	75.9
10	Learning to ride a bike	1	0.0
		2	0.0
		3	3.3
		4	82.8
11	Packing for a trip	1	0.0
		2	0.0
		3	6.7
		4	82.8
12	Brewing coffee	1	0.0
		2	0.0
		3	13.3
		4	79.3
13	Decorating the Christmas tree	1	0.0
		2	0.0
		3	13.3
		4	75.9
14	Eating an orange	1	0.0
		2	0.0
		3	13.3
		4	82.8
15	Playing the guitar	1	0.0
		2	0.0
		3	10.0
		4	72.4
16	Having a picnic	1	0.0
		2	0.0
		3	13.3
		4	82.8
17	Graduation	1	0.0
		2	0.0
		3	0.0
		4	82.8
18	Mouse in the house	1	0.0
		2	0.0
		3	3.3
		4	89.7
19	At a NASCAR race	1	0.0
		2	0.0
		3	6.7
		4	79.3

20	Listening to the car radio	1	0.0
		2	0.0
		3	10.0
		4	75.9
21	Getting a haircut	1	0.0
		2	0.0
		3	3.3
		4	86.2
22	Getting in a fight	1	3.4
		2	0.0
		3	6.7
		4	93.1
23	Gardening	1	0.0
		2	3.2
		3	10.0
		4	93.1
24	Opening the front door	1	3.4
		2	0.0
		3	6.7
		4	89.7
25	Teacher writing on a chalkboard	1	3.4
		2	0.0
		3	13.3
		4	79.3

Note. These discourses were retained based on the following title identification criteria: (a) < 14% accuracy for the Untitled groups after reading sentence 1 (Waves 1-2) and sentence 2 (Wave 3), (b) > 72% accuracy for the Untitled group after reading the complete 3-sentence long discourses (Wave 4).

APPENDIX C

COMPLETE SET OF 53 DISCOURSES USED FOR AMBIGUOUS NARRATIVE DISCOURSE ERP TASK

I. Experimental Discourses

1. Doing Laundry (adapted from Bransford & Johnson, 1972)
The man put the things into piles.
He then chose from among several buttons.
He put the first group in the washer.

2. Taking a driving test
The boy felt that he did not need more training.
He was sitting next to the instructor.
Then the boy started the car

3. Making a peanut butter and jelly sandwich
The man was expecting to make a small mess.
First, he found the necessary jars.
Then he had to get the bread.

4. **Going to the movies**
The man walked to the counter.
He made sure that he had his cash.
He already began to smell the popcorn.

5. Wedding ceremony
The man was in a room with many flowers.
He was surrounded by hundreds of attendants.
They were all watching the bride.

6. Eating ice cream
The man had spent the entire day outdoors.
When he got home he took out a bowl.
The only kind he had was vanilla.

7. Kids getting on the bus
The little girl heard the beep.
Then she quickly ran from her house.
She saw the other kids in the bus.

8. Stuck outside in the rain
The man quickly covered his head.
He knew he should have attended to the television.
He did not plan for this weather.

9. Gambling
The man wanted to get the prize.
He made sure not to adjust his sunglasses.
Unfortunately he did not have the right cards.

10. Learning to ride a bike
The boy would need more repetitions.
He had already gotten a few scratches.
He again pushed down on the pedals.

11. Packing for a trip
The woman went through her list.
She only had a limited amount of room.
She thought about getting a bigger suitcase.

12. Brewing Coffee
The woman eagerly waited in the kitchen.
The smell grew with every single droplet.
She took ahold of her favorite mug.

13. Decorating the Christmas tree
It almost did not fit in the doorway.
Then the dad went to get the ladder.
He still had to put on the star.

14. Eating an orange
It was easiest for the man to use his fingernails.
It almost instantly produced a smell.
He slowly began to take off the peel.

15. Playing the guitar
The woman found a comfortable grip.
Then she looked at the audience.
She began to move the strings.

16. Having a picnic
The couple could only do it during certain months.
They first had to find a large enough clearing.
The man first put down the blanket.

17. Graduation
All of the people had the same outfit.
Hundreds of others watched them from the bleachers.
The people waited to get their diplomas.

18. Mouse in the house
The family started on the hunt.
They were all looking for the intruder.
They would lure it in with some cheese.

19. At a NASCAR race
The event was filled with thousands of fans.
Some of the people protected their ears.
They could hardly see the cars.

20. Listening to the car radio
The couple was on a ride.
They continued to argue over the switch.
They could not decide on a station.

21. Getting a haircut
The woman looked through her calendar.
She wanted it done before the party.
All she needed was a trim.

22. Getting in a fight
The man had to use his arms.
He was prepared for the soreness.
He moved quickly to avoid the punch.

23. Gardening
The man made several holes.
Then he wiped off his jeans.
Next he put down some seeds.

24. Opening the front door
The girl stepped on to the mat.
Then she looked into her bag.
She needed to get her key.

25. Teacher writing on a chalkboard
The woman spoke at the wall.
She continued to rapidly make the marks.
She glanced back at the students.

II. Type 1 Fillers

26. Preparing eggs
She would first have to make an omelet.
The woman did it almost every morning.
Often times she could do it with one hand.

27. At the zoo
She loved to see all the animals.
The girl went with the other children.
She had to bring her own lunch.

28. Attacked by a bee
She was in discomfort from the sting.
The woman quickly began to get a rash.
She regretted that she forgot to bring pants.

29. Changing a diaper
She went to go get a new diaper.
The woman heard the cries.
She could not avoid the odor.

30. Eating pizza
Each boy waited for his slice.
The dad brought home the box.
All of his sons were at the table.

31. Trick-or-treating
Many of the people had on costumes.
It was rarely this busy in the neighborhood.
Occasionally there would be a loud scream.

32. Brushing teeth
She wanted to please her dentist.
The woman first removed the cap.
There was not much left in the tube.

III. Type 2 Fillers

33. Power outage
It affected everyone on the block.
Their daily routines relied on the electricity.
The family was not able to use the shower.

34. Studying
The boy just wanted a break.
He was carrying all of his books.
He was expecting it to be another difficulty night.

35. Running
The man stayed on the asphalt.
He planned to do several more miles.
His body continued on in the same pattern.

36. Bowling
The lights reflected off of the wood.
She tried to get all of the pins.
The girl closely watched the score.

37. Washing dishes
The woman did not have the necessary machine.
She picked up the first plate.
She did not have any help this evening.

38. Recycling
It did not work for all products.
It was his way of helping the environment.
The man kept everything in separate containers.

39. Singing karaoke
The man only did it after several drinks.
He hoped that he knew the words.
He slowly made his way to the front.

IV. Type 3 Fillers

40. Birthday Party
The little girl stared at her mom's watch.
All of a sudden she heard the doorbell.
Everyone walked in with a sky.

41. Fishing
The man had started right after sunrise.
He almost did not notice the ripple.
He had to quickly grab his cloud.

42. A day at the beach
The family came from the other side of the mountains.
They were thankful that there was no rain.
They were all walking on the broom.

43. Sitting at a bonfire
The friends were mesmerized by the glow.
Everything else was covered in darkness.
They moved away to avoid the yogurt.

44. At the circus
This only occurred during the summer.
There was always an extensive line.
Many people were inside the coin.

45. Learning to swim
The kids attended to the teacher.
She showed them all the motions.
To start they took a large magic.

46. Mailing a letter
It would have been quicker to use a phone.
The woman finished with her name.
She checked her accuracy on the front of the gum.

V. Type 4 Fillers

47. Playing with a dog
The man was holding his friend's leash.
The man and his friend waited on the BENCH.
The man made sure to bring several games.

48. Cooking
She only needed a few ingredients.
The girl read the words on the PACKAGE.
Then she took out the necessary tools.

49. Using the mirror
She stepped to the mirror.
The woman only used it for one JOB.
She decided to change her shoes.

50. Picture day at school
She looked at the camera.
The girl sat in front of big LAMPS.
Then she quickly adjusted her shirt.

51. Making pancakes
One at a time he took them off the griddle.
The man patiently watched the CIRCLES.
Soon they would all be in his mouth.

52. Singing the national anthem
They were all staring at the flag.
It was just before the WHISTLE.
Everyone was waiting for the leader.

53. Shaving
He began to use the razor.
The man rubbed his FACE.
Then he slightly raised his chin.