

An investigation of concurrent ERP and self-paced reading methodologies

TALI DITMAN,^a PHILLIP J. HOLCOMB,^a AND GINA R. KUPERBERG^{a,b}

^aDepartment of Psychology, Tufts University, Medford, Massachusetts, USA

^bDepartment of Psychiatry and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, USA

Abstract

Traditionally, event-related brain potential (ERP) studies of language processing have presented words at a fixed rate using rapid serial visual presentation. Recent studies suggest, however, that the processes engaged during sentence comprehension are contingent on word presentation rate. These findings underscore the importance of allowing participants to read at a natural pace. The present study employed simultaneous self-paced reading and ERP methodologies to examine behavioral and neural responses while participants read sentences containing pragmatic or morphosyntactic violations or no violations. ERP and self-paced reading results replicated previous findings. This novel combination of behavioral and ERP methodologies combines the high temporal resolution and direct neural measures offered by ERPs with the more natural reading environment and information about processing load provided by self-paced reading.

Descriptors: Event-related potentials, Self-paced reading, P600, N400, Language, Sentence, Methodology

Reading comprehension is a complex process and, as such, has been examined using a variety of techniques, such as self-paced reading, eye tracking, probe verification tasks, event-related brain potentials (ERPs), and neuroimaging techniques. Although each methodology provides useful information, no research paradigm is without its drawbacks (for a discussion, see Haberlandt, 1994). These techniques vary in their ability to capture a natural reading environment as well as in the type of information they provide (e.g., immediate lexical and/or integrative postlexical and qualitative and/or quantitative information). Thus, the most comprehensive understanding of language comprehension can only be gleaned through the combination of multiple techniques.

The present article focuses on the combination of ERP and self-paced reading methodologies. ERPs are derived from the ongoing electroencephalogram (EEG), which is a measure of the electrical field resulting from the firing of many neurons (10^3 – 10^4 cells) across time. They provide important insights into the neurocognitive processes engaged during reading comprehension

(for a description of ERPs, see Kutas & Van Petten, 1994; Rugg & Coles, 1995). ERPs are obtained by time-locking electrical activity to a specific point of interest (e.g., a critical word). They allow for a direct online assessment of neural activity with millisecond temporal resolution. In addition, ERPs allow for a qualitative analysis of the data as well as a quantitative analysis. Thus, they give insights not only into *when* a difference between conditions occurs but also into the type of neurocognitive process that this difference reflects. This is in contrast to other methodologies, such as self-paced reading techniques, in which increases/decreases in average reading times between conditions provide evidence for quantitative differences, but cannot provide information about the processes involved in creating these differences (for a similar discussion, see Haberlandt, 1994; McKoon & Ratcliff, 1980).

Traditionally, ERP language research has used a fixed-rate rapid serial visual presentation (RSVP) of words. Using this methodology, words are presented serially in the center of a computer screen for a set amount of time (e.g., 400 ms) after which the word disappears and is replaced by another word. As with all techniques, fixed RSVP has both advantages and disadvantages. An important advantage of fixed RSVP is that the researcher is able to control the duration of word presentation. Thus, all participants have the same amount of time to process words in all conditions, ensuring that differences between conditions are not the result of processing time differences. In addition, in ERP research, this ensures that all early components (e.g., the N1-P2 complex) to upcoming words will be occurring at specific intervals. Thus, one can time-lock neural activity to a particular word and examine activation to subsequent words.

We are grateful to Tad Brunye, Evelina Fedorenko, Ted Gibson, and Mante Nieuwland for their comments and insights. This research was supported by NIMH (R01 MH071635 to Gina R. Kuperberg). Gina R. Kuperberg was also supported by NARSAD (with the Sidney Baer Trust) and by a Claflin Distinguished Scholars Award from Massachusetts General Hospital. This research was also supported by NICHD grants HD25889 and HD043251 to Phillip J. Holcomb and NIDCD grant DC05237 to Neal Pearlmuter.

Address reprint requests to: Tali Ditman, Tufts University, Psychology Building, 490 Boston Ave., Medford, MA 02155, USA. E-mail: tali.ditman@tufts.edu

A disadvantage, however, of a fixed RSVP is that it is not a natural way to read and, as such, results may not necessarily reflect processes that occur during normal reading comprehension. To address this issue, some researchers have attempted to increase naturalness by varying the rate of word presentation by the number of characters in a word (Nieuwland & Van Berkum, 2006; for a similar variable presentation rate in behavioral language studies, see Haberlandt & Graesser, 1985; Legge, Ahn, Klitz, & Luebker, 1997). This technique is advantageous in comparison to fixed RSVP as it better simulates the conditions influencing natural reading times, allowing increased reading comfort for participants. However, this technique is not able to address reading speed differences between individuals.

Another disadvantage of using the fixed-rate RSVP display is that different presentation rates may bias toward the engagement of different cognitive processes. Evidence for this comes from a recent set of studies suggesting that faster presentation rates are associated with an increased influence of lexical factors, and slower rates are associated with an increased influence of higher level (i.e., sentence- and discourse-level) factors on the modulation of ERPs (Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Ledoux, Gordon, Camblin, & Swaab, in press; Swaab, Camblin, & Gordon, 2004). For example, in a recent ERP investigation of lexico-semantic and sentence-level influences on word recognition, Swaab et al. found an immediate influence of sentence context rather than lexico-semantic influences when an SOA of 700 ms was employed (500 ms/word and 200 ms interstimulus interval [ISI]). This artificial presentation rate, however, may have provided time for the lexico-semantic influences to fade and the effects of discourse influences to appear. These results were replicated by Ledoux and colleagues, using a stimulus onset asynchrony (SOA) of 500 ms (300 ms/word and 200 ms ISI). Interestingly however, using a fast presentation rate (300 ms/word and 50 ms ISI), immediate lexico-semantic rather than sentence-level effects were observed (Camblin et al.). Moreover, when an eye-tracking methodology, in which participants advanced to the next word at their own pace, was used, results replicated the effects of the fast rather than the slow presentation rate, although effects of sentence context were observed at a later processing point (Ledoux et al.). The diverging results of eye-tracking, slow-rate RSVP, and fast-rate RSVP underscore the importance of employing a methodology that is sensitive to online language comprehension processes, that provides both quantitative and qualitative information, but that still allows readers to read at a natural and comfortable pace for comprehension.

One potential solution to the presentation rate problem is to record ERPs as participants are allowed to advance through sentences word by word at their own pace. In addition to allowing for a more natural presentation rate, coupling ERPs with self-paced reading methodology provides an opportunity for direct examination of the relationship between evoked brain potentials and reading times. Although previous research has compared ERPs and reading times in separate experiments, differential task demands preclude any direct comparisons. In addition, separate ERP and self-paced reading experiments do not allow one to assess potential relationships between the two dependent measures. Specifically, by simultaneously collecting both reading time and ERP measures from the same participants, one can perform direct statistical analyses to elucidate any potential differences and similarities.

Despite these advantages, to our knowledge this dual methodology has not previously been employed. This is because it has usually been assumed that asking participants to make a motor response to each word would increase the amount of EEG artifact due to motor responses and decision making, thereby limiting the sensitivity of ERPs to detect differences between conditions. This assumption, however, has not yet been verified experimentally. Importantly, if a combination of these techniques is viable, it may increase our ability to detect processes that occur during normal reading comprehension as well as participants' subjective experience of naturalness during reading comprehension in an ERP experiment relative to traditional RSVP.

To examine the feasibility of using simultaneous self-paced reading and ERP techniques, we used a well-established paradigm and asked participants to read sentences that contained either pragmatic¹ or morphosyntactic violations (Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006). Violations occurred midsentence, on the verb.

Consistent with numerous other studies introducing semantic/pragmatic violations within sentences (De Vincenzi et al., 2003;² Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Kuperberg et al., 2006; Kutas & Hillyard, 1980, 1984), Kuperberg et al. (2003, 2006) found that, relative to morphosyntactic violations or nonviolated words, semantic/pragmatic violations elicited a larger amplitude waveform from 300 to 500 ms (peaking at approximately 425 ms) with a centroparietal distribution. This waveform, termed the N400, is thought to reflect the ease of semantically integrating a word into its preceding context (e.g., Holcomb, 1993; for a review, see Kutas & Federmeier, 2000).

In addition, consistent with other studies introducing morphosyntactic violations in sentences (e.g., De Vincenzi et al., 2003; Hagoort, 2003; Kuperberg et al., 2003), Kuperberg et al. (2006) found that, relative to the nonviolated and semantically/pragmatically violated verbs, the morphosyntactic violations elicited a larger amplitude positive-going waveform³ from 500 to 900 ms, also with a centroparietal distribution. This waveform is termed the P600 (Osterhout & Holcomb, 1992). Although the precise neurocognitive processes reflected by the P600 are debated (syntactic integration, monitoring, or reanalysis), it is acknowledged to be highly sensitive to multiple types of

¹We follow Marslen-Wilson, Brown, and Tyler (1988) in the use of the term "pragmatic" for these types of violations. These sentences all describe possible scenarios that are less plausible with respect to our real-world knowledge relative to the nonviolated sentences. Note, however, our use of the term "pragmatic" is distinct from the use of this term to refer to the relationship between sentence meaning and speaker's meaning and/or the use of the term to refer to a variety of phenomena at the level of discourse.

²De Vincenzi et al. refer to these violations as being selection restriction violations. Examination of their stimuli, however, reveals the use of several different types of meaning violations on the verb, including some conferred by pragmatic real-world knowledge, some that are unpredictable with respect to the preceding context, some that are conferred by selection restriction constraints between the verb and its subject NP argument, and some conferred by an unusual use of verbs. In this article, we therefore use the umbrella term "semantic" to refer to these violations.

³Although there is some debate concerning whether morphosyntactic violations also elicit a left anterior negativity (LAN), as found in some studies (Osterhout & Mobley, 1995) but not others (Gunter & Friederici, 1999; Hagoort, Brown, & Groothusen, 1993; Osterhout & Nicol, 1999), we did not predict an LAN in the present study as two previous ERP studies using these stimuli with different participants did not find this effect (Kuperberg et al., 2003, 2006).

syntactic violations and ambiguities (for a review see Osterhout & Holcomb, 1995).

Previous behavioral studies have also contrasted the effects of introducing semantic and syntactic anomalies on reading times during self-paced reading paradigms. De Vincenzi et al. (2003) presented participants with Italian sentences containing mid-sentence morphosyntactic violations, consisting of subject–verb agreement errors, or semantic violations, consisting of content violations on mid-sentence verbs. Results demonstrated that, in comparison with critical words in nonviolated sentences, morphosyntactic violations led to increased reading times at the point of the anomaly. Reading times at the sentence-final word, however, were similar in both the morphosyntactically violated and nonviolated sentences. On the other hand, in comparison within nonviolated sentences, semantic violations did not lead to any increase in processing load at the point of the anomaly, but rather led to longer reading times to several words following the critical word as well as on the sentence-final word.

The aim of the current study was to investigate the feasibility of using ERP and self-paced reading in the same participants at the same time by determining whether these previous findings could be replicated using this dual methodology.

METHODS

Participants

Eighteen native English speakers (9 male/9 female, average age 19.7 years) participated for monetary compensation. Written informed consent was obtained from all participants in accordance with the guidelines of the Tufts Human Subjects Research Committee. All participants were right-handed, had no history of traumatic head injury, and had normal or corrected-to-normal vision.

Design and Materials

The materials have been described in detail by Kuperberg et al. (2006).⁴ Participants read 180 sentences in total, 60 that contained a pragmatic violation, 60 that contained a morphosyntactic violation, and 60 with no violations (sentence type: pragmatic violation, morphosyntactic violation, nonviolated). Briefly, sentences with pragmatic violations were constructed by replacing the critical verb of each nonviolated sentence (e.g., *At breakfast the boys would eat toast and jam.*) with verbs that were chosen pseudorandomly from sentences from another list (e.g., *At breakfast the boys would plant toast and jam.*). Sentences with morphosyntactic violations were created by either violating subject–verb agreement or by using a finite in place of an infinitival verb (e.g., *At breakfast the boys would eats toast and jam.*). The violation always occurred on a critical verb that was never the sentence-final word. On average, sentences were 11.2 words long, with 7.9 words prior to the critical word and 2.3 words following the critical word.

Sentences were counterbalanced such that the same critical verb appeared in different conditions on each of three lists. Each participant only saw one list and thus only read each critical word once. Within each list, the order of sentence presentation was randomized.

⁴The present study did not use the animacy violations that had been employed in Kuperberg et al. (2006).

Procedure

Participants were seated in front of a computer monitor and were instructed to read all sentences at their own pace for comprehension. Each trial began with the word “READY” in capital letters in the center of the screen during which participants were able to blink. When participants were ready to begin the trial, they pressed a button on a game-pad and a fixation cross appeared in the center of the screen. Hand position was counterbalanced such that half of the participants advanced to the next word using their left thumb and the other half of participants used their right thumb. Upon pressing the game-pad button to advance, the first word of the sentence appeared centered on the screen. With each button press, the sentence unfolded in the center of the screen word by word. There was a 700-ms ISI between the last word of the sentence and a “?”. At the question mark, participants were asked to make an acceptability judgment about the sentence. Participants were told that sentences should be deemed unacceptable either if the sentence contained a grammatical error or if the content of the sentence was odd. Each participant was given six practice trials before beginning the experiment.

Recording Procedure

Participants were seated in a comfortable chair in a sound-attenuated darkened room. An elastic cap (Electro-Cap International) with 29 active tin electrodes was placed on the participant’s head. The electrodes were located in the standard International 10–20 System locations as well as at additional sites over the left and right hemispheres (see Figure 1). Electrode locations consisted of five sites along the midline (FPz, Fz, Cz, Pz, Oz), three medial electrode sites over each hemisphere (FC1/FC2, C3/C4, CP1/CP2), four lateral electrodes over each hemisphere (F3/F4, FC5/FC6, CP5/CP6, P3/P4), and five peripheral sites over each hemisphere (FP1/FP2, F7/F8, T3/T4, T5/T6, O1/O2). To monitor vertical eye movements/blinks and horizontal eye movements, electrodes were placed below the left eye and lateral to the right eye, respectively. Electrodes were referenced to the left mastoid and an electrode was placed on the right mastoid to monitor differential mastoid activity.

The EEG was amplified by a SA Bioamplifier using a band-pass of 0.01–40 Hz and was continuously sampled at a rate of 200 Hz. Electrode impedances were kept below 10 k Ω for the eyes and below 5 k Ω at all other sites. For each participant, separate ERPs were averaged off-line at each electrode site for each experimental condition. Trials contaminated with eye artifact or amplifier blockage were not included in analyses.

Behavioral Data Analysis

Mean reading times to the critical word and the sentence-final word of each sentence were entered into separate one-way (sentence type: pragmatic violation, syntactic violation, no violation) repeated measures analyses of variance (ANOVAs). Reading times were analyzed by subjects (F1) and by items (F2). For consistency across behavioral and ERP data, we report data only from trials that were not artifact rejected in the ERP analysis.⁵ In

⁵Analyses were also performed on trials on which participants responded correctly to the probe as well as on all trials regardless of response. Additional analyses were conducted to adjust for differences in word lengths and differences in participants’ reading rates. A regression equation was calculated to predict reading times from word length for each participant across all items (for a discussion, see Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994). Specifically, for each

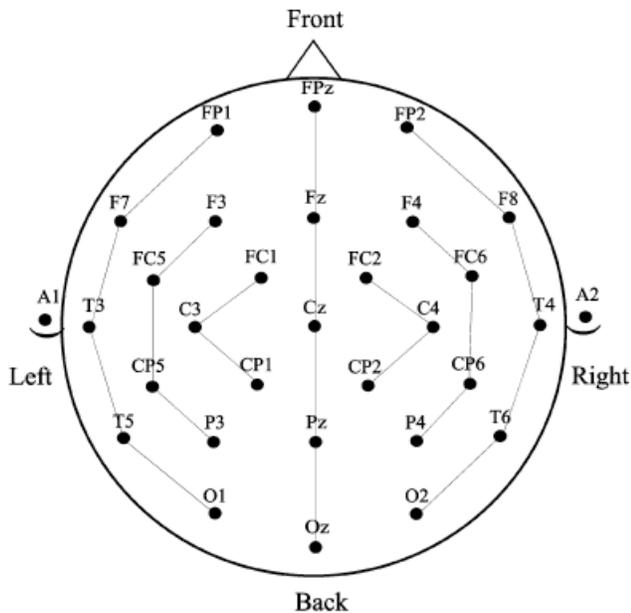


Figure 1. Electrode montage. Electrodes placed in the standard International 10–20 System locations included five sites along the midline (FPz, Fz, Cz, Pz, and Oz) and eight lateral sites, four over each hemisphere (F3/F4, C3/C4, T3/T4, and P3/P4). Eight additional 10–20 sites were altered to form a circle around the perimeter of the scalp. These altered sites included FP1'/FP2' (33% of the distance along the circle between T3/T4), F7'/F8' (67% of the distance between FPz and T3/T4), T5'/T6' (33% of the distance between T3/T4 and Oz), and O1'/O2' (67% of the distance between T3/T4 and Oz). In addition eight extended 10–20 system sites were also used (FC1/FC2, FC5/FC6, CP1/CP2, and CP5/CP6). The dotted lines represent the four columns used in analyses (i.e., midline, medial, lateral, peripheral).

addition, to examine performance on the acceptability judgment task, accuracy and response time (RT) data were analyzed by means of separate one-way ANOVAs. Paired *t* tests were conducted to investigate significant interactions.

ERP Data Analysis

All analyses were conducted on mean amplitude values using the 100 ms of activity that preceded word onset as a baseline. Analyses were conducted on all trials, regardless of accuracy.⁶ Two time windows were chosen for examination: 300–500 ms (corresponding to the N400) and 500–1000 ms (corresponding to the P600). ERPs in both these time windows were examined to both the critical verbs and the sentence-final words. Repeated measures ANOVAs were performed on the midline, medial, lateral, and peripheral sites described above. In addition, in all analyses Sentence Type was entered as a within-subject factor and, for the medial, lateral, and peripheral analyses, Hemisphere (left, right) was an additional within-subjects factor. A Greenhouse–Geisser correction was applied to all analyses with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959). In these cases, the original degrees of freedom with the corrected

p value are reported. Significant interactions were further examined with simple effects tests.

Regression Analyses

To determine whether, across items, ERP effects on the critical word predicted reading times, two regression analyses were conducted: (1) N400 amplitude evoked to each critical word at Pz was regressed to the reading time for the corresponding item and (2) P600 amplitude evoked to each critical word at Pz was regressed to the reading time for the corresponding item. In addition, as we predicted largest reading time differences to pragmatic violations at the sentence-final word based on previous studies (e.g., for similar findings using other semantic violations, see De Vincenzi et al., 2003), these same regressions were performed using reading times to sentence-final words.⁷

Examining Potential Neural Contamination Due to Reading Time Differences

To examine whether systematic differences in the latency of early components to subsequent words in the different conditions may have confounded N400 and P600 effects, we subtracted early neural differences to the word following the critical word from N400 and P600 effects at electrode site Pz. Specifically, if the mean reading time for critical words in a specific condition was 450 ms, then exogenous components could have influenced the amplitude of this wave between 450 and 500 ms following critical word onset. To account for this, we subtracted mean neural activity between 0 and 50 ms to the word following the critical word from the mean amplitude of the critical word between 450 and 500 ms. Using this adjusted value, we then computed a new mean amplitude for the 300–500-ms time window. In this way, for each condition, the N400 amplitude was adjusted to account for latency differences in incoming early components to subsequent words.

A similar calculation was conducted on the P600 amplitude between 500 and 1000 ms following critical word onset. Specifically, using a similar example, if the mean reading time for a specific condition was 450 ms and we define early components as differences between 0 and 250 ms following word onset, then early components to the subsequent word could have influenced the P600 amplitude between 500 and 700 ms. To account for this, mean neural activity to the word following the critical word between 50 and 250 ms was subtracted from mean amplitude differences between neural activity evoked to the critical word at 500–700 ms. In this way, for each condition, the P600 amplitude was adjusted to account for latency differences in incoming early components to subsequent words.

Results

Behavioral Data (see Figure 2)

Critical verb. Participants were slowest to read morphosyntactic violations ($M = 451.41$ ms, $SD = 187.43$) compared with both pragmatic violations ($M = 416.71$ ms, $SD = 139.21$) and nonviolated ($M = 407.47$ ms, $SD = 136.14$) critical words. Repeated measures ANOVAs confirmed this difference: $F(1, 2,34) = 6.31$, $p < .05$; $F(2, 358) = 10.53$, $p < .001$, and

item, the reading time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. Identical patterns of results were obtained in all analyses.

⁶Additional analyses restricted to accurate trials yielded identical patterns.

⁷Regression analyses were not conducted on the word following the critical word, a point at which spillover effects are sometimes observed in reading time studies, as no reading time differences were observed between the three violations at this point, $F(2, 34) = 1.97$, $p > .10$.

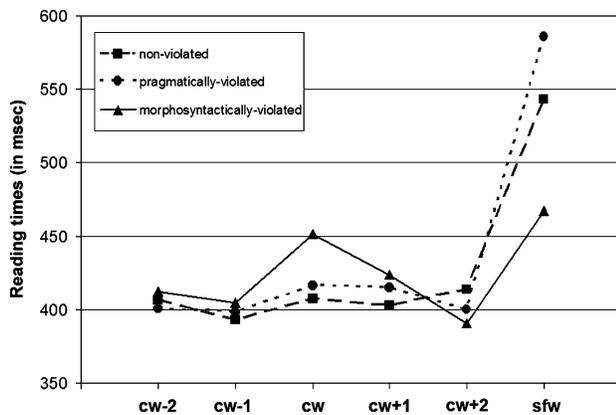


Figure 2. Reading times to the three sentence types beginning two words prior to the critical word (cw - 2), extending to two words after the critical word (cw + 2), and on the sentence-final word (sfw).

paired *t* tests demonstrated that, whereas participants were slower to read morphosyntactic violations than pragmatic violations ($ps < .05$) and nonviolated ($ps < .05$) critical words, reading times to pragmatic violations and nonviolated verbs did not differ from one another ($ps > .10$).

Sentence-final word. The pattern of results changed at the sentence-final word. Participants were fastest to read final words of sentences that contained morphosyntactic violations ($M = 467.32$ ms, $SD = 158.59$) compared with sentences that contained pragmatic violations ($M = 586.03$ ms, $SD = 248.65$) or nonviolated ($M = 543.33$ ms, $SD = 194.33$) sentences. This difference was confirmed by ANOVAs: F1: $F(2,34) = 8.30$, $p < .01$; F2: $F(2,358) = 17.50$, $p < .001$, and paired *t* tests showed that final words of sentences containing morphosyntactic violations were read faster than final words of the other two sentence types (all $ps < .05$). Additionally, final words of sentences containing pragmatic violations were read slowest, although the difference between pragmatic and nonviolated conditions was only marginally significant in the subjects analysis ($p < .10$) but significant in the items analysis ($p < .05$).

Acceptability judgment performance. Participants were both more accurate and faster at judging sentences containing morphosyntactic violations as unacceptable (accuracy: $M = 97.41\%$, $SD = 2.69$; RT: $M = 616.84$ ms, $SD = 318.52$) than judging sentences containing pragmatic violations as unacceptable (accuracy: $M = 88.52\%$, $SD = 7.34$; RT: $M = 851.07$ ms, $SD = 433.18$) or nonviolated sentences as acceptable (accuracy: $M = 91.48\%$, $SD = 6.15$; RT: $M = 860.25$ ms, $SD = 418.76$). These results were confirmed by main effects of Sentence Type for both the accuracy, $F(2,34) = 10.12$, $p < .01$, and RT data, $F(2,34) = 9.76$, $p < .01$. Paired *t* tests demonstrated that, in both accuracy and RT analyses, morphosyntactic violations differed from the other two sentence types (all $ps < .01$), which did not differ from one another (all $ps > .10$).

ERP Data

Artifact contamination from eye movement or amplifier blocking led to the rejection of 12.1% of the trials at the critical word and 16.6% of the trials at the sentence final word. The number of

rejected items did not differ by condition ($ps > .05$ for all pairwise comparisons).

Figures 2 (ERP waves to the critical word) and 3 (ERP waves to the sentence-final word) show a negative wave component occurring from 100 to 200 ms and a positive component from 200 to 300 ms following word onset (the N1 and P2 components). The N1-P2 complex was followed by a negative-going component (i.e., the N400) between 300 and 500 ms. At the critical word (Figure 3) the N400 was followed by a positive deflection from 500 to 1000 ms (i.e., the P600).

Critical Word

300–500 ms. As displayed in Figure 3, within 300–500 ms after critical word onset, a large negativity, the N400, was evoked to pragmatic violations compared with morphosyntactic violations and nonviolated words. Main effects of Sentence Type demonstrated that the N400 to the three conditions differed from one another: midline: $F(2,34) = 17.27$, $p < .001$; medial: $F(2,34) = 15.27$, $p < .001$; lateral: $F(2,34) = 12.86$, $p < .001$; peripheral: $F(2,34) = 11.80$, $p < .001$. Specifically, whereas the N400 to morphosyntactic and nonviolated critical words did not differ (pairwise comparisons at midline, medial, lateral, and peripheral electrode columns: $p > .10$), the N400 amplitude to pragmatic violations was significantly larger than the N400 amplitude to morphosyntactic violations (pairwise comparisons at midline, medial, lateral, and peripheral electrode columns: $p < .001$) or nonviolated words (pairwise comparisons at all columns: $p < .01$).

This effect appeared all over the scalp but, typical of the N400 scalp distribution, it was largest at centroparietal sites, as evidenced by Sentence Type \times Electrode Sites interactions at all columns: midline: $F(8,136) = 7.48$, $p < .001$; medial: $F(4,68) = 9.93$, $p < .001$; lateral: $F(6,102) = 9.49$, $p < .001$; peripheral: $F(8,136) = 5.37$, $p < .01$. Follow-up simple effects ANOVAs revealed main effects of Sentence Type at all but three anterior sites (midline: FPz: $p < .10$; peripheral: FP1/2: $p > .10$; F7/8: $p > .10$). This effect was slightly larger on the right hemisphere, demonstrated by Sentence Type \times Hemisphere interactions at medial, lateral, and peripheral columns: medial: $F(2,34) = 3.81$, $p < .05$; lateral: $F(2,34) = 6.98$, $p < .01$; peripheral: $F(2,34) = 9.70$, $p < .001$. Follow-up simple effects ANOVAs revealed significant main effects of Sentence Type on both the left (all $F_s > 4.04$, all $ps < .05$) and the right (all $F_s > 16.65$, all $ps < .001$) hemispheres. Furthermore, at medial and lateral columns, a Sentence Type \times Electrode Sites \times Hemisphere interaction was also observed: medial: $F(4,68) = 4.73$, $p < .01$; lateral: $F(6,102) = 3.19$, $p < .05$; peripheral: $F(8,136) = 1.74$, $p > .10$.

500–1000 ms. A large positive deflection, the P600, was observed to morphosyntactic violations compared to pragmatic violations and nonviolated critical words (see Figure 3), demonstrated by main effects of Sentence Type: midline: $F(2,34) = 31.05$, $p < .001$; medial: $F(2,34) = 32.00$, $p < .001$; lateral: $F(2,34) = 25.36$, $p < .001$; peripheral: $F(2,34) = 21.88$, $p < .001$. Morphosyntactic violations evoked a larger P600 than pragmatic violations (all pairwise comparisons, $ps < .001$) and nonviolated critical words (all pairwise comparisons, $ps < .001$), which did not differ from one another (all pairwise comparisons, $ps > .10$).

The scalp distribution of this effect was larger at right centroparietal sites as demonstrated by Sentence Type \times Electrode Sites interactions at all columns: midline: $F(8,156) = 28.55$,

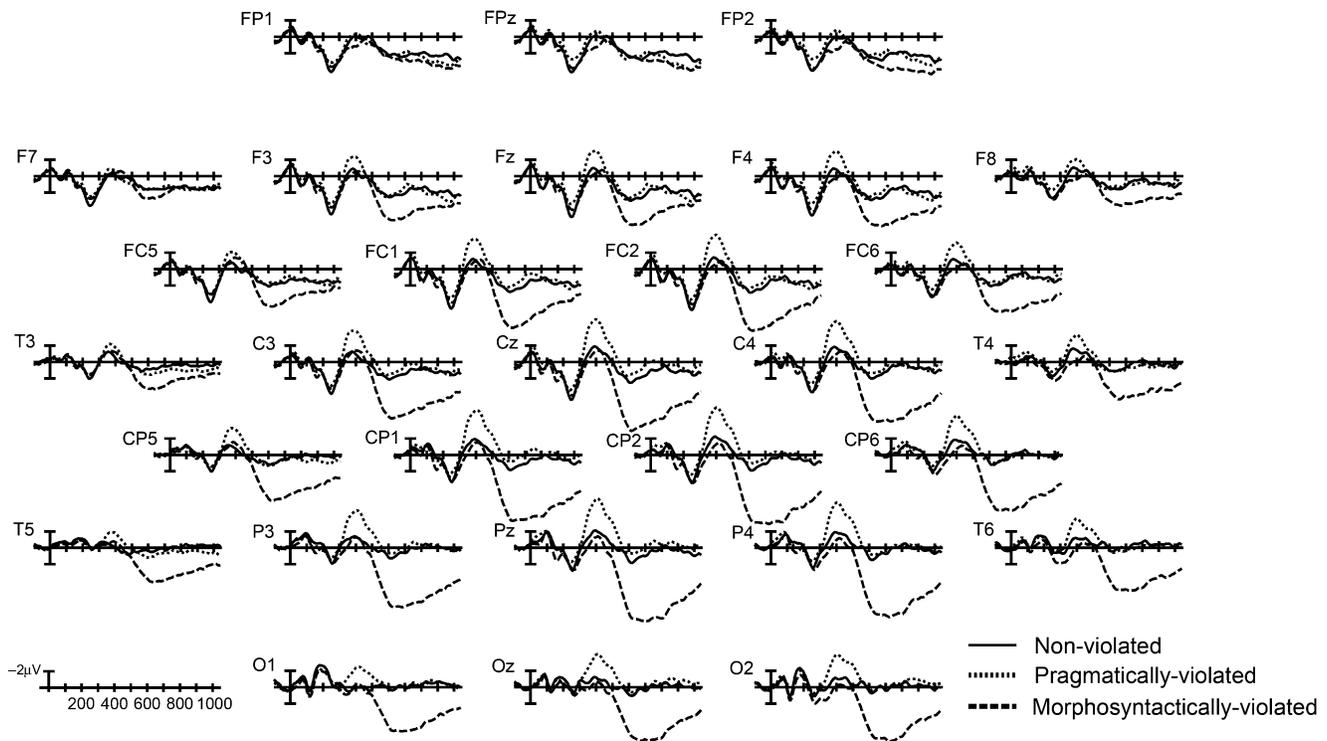


Figure 3. ERPs at the critical word to the three sentence types. The solid line shows ERPs to critical words in nonviolated sentences, the dotted line shows ERPs to critical words in sentences with pragmatic violations, and the dashed line shows ERPs to critical words in sentences with morphosyntactic violations.

$p < .001$; medial: $F(4,68) = 30.38$, $p < .001$; lateral: $F(6,102) = 33.76$, $p < .001$; peripheral: $F(8,136) = 18.53$, $p < .001$, and Sentence Type \times Hemisphere interactions at lateral and peripheral columns: medial: $F(2,34) = 2.34$, $p > .10$; lateral: $F(2,34) = 3.90$, $p < .05$; peripheral: $F(2,34) = 5.61$, $p < .01$. The interaction of Sentence Type \times Electrode Sites \times Hemisphere was not significant at any column: medial: $F(4,68) = 2.38$, $p < .10$; lateral: $F(6,102) = 2.17$, $p < .10$; peripheral: $F(8,136) = 1.33$, $p > .10$).

Sentence-Final Word

300–500 ms. Figure 4 depicts the largest negativity to final words of sentences containing morphosyntactic violations, followed by a medium-sized N400 to sentences containing pragmatic violations and the smallest amplitude N400 to nonviolated sentences. Main effects of Sentence Type were observed at all columns: midline: $F(2,34) = 20.21$, $p < .001$; medial: $F(2,34) = 19.09$, $p < .001$; lateral: $F(2,34) = 16.90$, $p < .001$; peripheral: $F(2,34) = 13.97$, $p < .001$, demonstrating that all three sentence types differed from one another (all pairwise comparisons, $ps < .01$). This difference was most pronounced at centroparietal sites, verified by Sentence Type \times Electrode Sites interactions at all columns: midline: $F(4,136) = 8.09$, $p < .001$; medial: $F(4,68) = 4.84$, $p < .05$; lateral: $F(6,102) = 10.21$, $p < .001$; peripheral: $F(8,136) = 7.53$, $p < .01$. Follow-up simple effects ANOVAs revealed main effects of Sentence Type at all sites (all $F_s > 7.70$, all $ps < .05$) with the exception of three anterior sites (midline: FPz: $p < .10$; peripheral: FP1/2: $p > .10$, F7/8: $p > .10$). There were no hemispheric differences (no Sentence Type \times Hemisphere interactions, all $F_s < 1.60$, all $ps > .10$, and

no Sentence Type \times Electrode Sites \times Hemisphere interaction, all $F_s < 1$, all $ps > .10$).

500–1000 ms. As can be seen in Figure 4, whereas the negativity to final words of sentences containing pragmatic violations returned to a similar level of activation as final words of nonviolated sentences, the negativity was sustained for final words of sentences containing morphosyntactic violations. Main effects of Sentence Type were observed: midline: $F(2,34) = 16.72$, $p < .001$; medial: $F(2,34) = 16.84$, $p < .001$; lateral: $F(2,34) = 15.72$, $p < .001$; peripheral: $F(2,34) = 13.60$, $p < .001$, but unlike the 300–500-ms time window, only final words of morphosyntactic violations had larger amplitude negativity than the other two sentence type conditions (all $ps < .01$), which did not differ from one another (all $ps > .10$).

The negativity was again maximal at centroparietal sites, evidenced by Sentence Type \times Electrode Sites interactions at all columns: midline: $F(8,136) = 9.10$, $p < .001$; medial: $F(4,68) = 7.02$, $p < .01$; lateral: $F(6,102) = 13.01$, $p < .001$; peripheral: $F(8,136) = 7.93$, $p < .001$, reaching significance at all but four anterior sites (midline: FPz: $p < .10$; medial: FC1/2: $p < .10$; peripheral: FP1/2: $p < .10$, F7/8: $p < .10$). Sentence Type \times Hemisphere interactions demonstrated that this negativity was maximal over the right hemisphere: medial: $F(2,34) = 5.07$, $p < .05$; lateral: $F(2,34) = 4.98$, $p < .05$; peripheral: $F(2,34) = 3.30$, $p = .05$, although follow-up simple effects ANOVAs showed that this effect was present on both left (all $F_s > 9.18$, all $ps < .01$) and right (all $F_s > 17.50$, all $ps < .001$) hemispheres. Finally, an interaction of Sentence Type \times Electrode Sites \times Hemisphere was only observed at the medial column: medial:

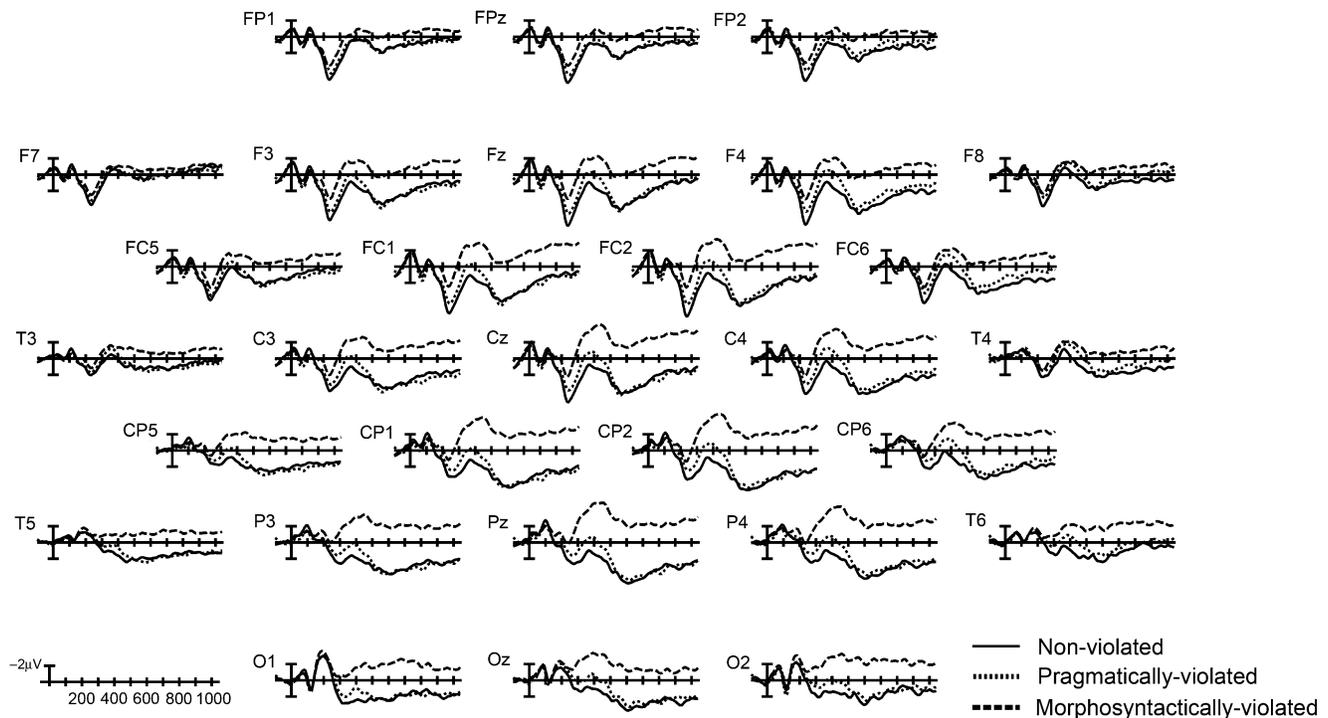


Figure 4. ERPs at the sentence-final word to the three sentence types. The solid line shows ERPs to final words in nonviolated sentences, the dotted line shows ERPs to final words in sentences with pragmatic violations, and the dashed line shows ERPs to final words in sentences with morphosyntactic violations.

$F(4,68) = 2.91, p < .05$; lateral: $F(6,102) < 1, p > .10$; peripheral: $F(8,136) < 1, p > .10$.

Regression Analyses

N400 amplitude evoked to critical words did not predict reading time to those same critical words, $\beta = .002, t(178) = 1.40, p > .10$. Interestingly, however, N400 amplitude to the critical word did predict reading time to sentence-final words, $\beta = -.011, t(178) = -3.29, p < .001$, with larger N400 amplitude predicting longer reading time. P600 amplitude evoked to critical words predicted reading time to critical words, $\beta = .004, t(178) = 3.905, p < .001$, as well as sentence-final words, $\beta = -.009, t(178) = -4.29, p < .001$. Specifically, larger P600 amplitude to the critical word predicted longer reading time to critical words but shorter reading time to sentence-final words.

Examining Potential Neural Contamination Due to Reading Time Differences

Figure 5 depicts mean N400 and P600 amplitudes to each sentence type at electrode site Pz before adjusting for early component latency differences to subsequent words and after adjusting for these latency differences. Differences between the conditions remained the same both visually and statistically before and after adjusting for early component latency differences. Main effects of Sentence Type replicated the N400, $F(2,34) = 23.24, p < .001$, and P600 effects, $F(2,34) = 62.66, p < .001$, described above.

Discussion

The present study assessed the viability of a novel methodology in the study of language comprehension: simultaneous self-paced reading and ERP techniques.

Despite potential concerns that this dual methodology would lead to increased artifact in the EEG signal, the results of the current study replicated both previous ERP and behavioral findings: Pragmatic violations evoked a robust N400 effect and syntactic violations evoked a robust P600 effect on the midsentence critical verbs (for similar findings using these stimuli, see Kuperberg et al., 2003, 2006). Sentence-final (noncritical) words in the pragmatically violated sentences also evoked an N400 effect (for similar findings, see Hagoort, 2003), and sentence-final (noncritical) words in the syntactically violated sentences evoked a sustained negativity effect (for similar findings, see Hagoort, 2003; Osterhout & Holcomb, 1992).⁸ Behaviorally, longer reading times were found immediately at the point of the morphosyntactically violated critical verbs, but later, on the sentence-final word, in pragmatically violated sentences (for similar findings, see De Vincenzi et al. 2003).

Importantly, in the present study, we did not find that increased artifact due to preparatory or motor responses associated with advancing from word to word obscured our ability to detect differences between conditions. There are several potential reasons for this. First, the button press used to advance to the next word was consistent across all conditions. Thus, any artifacts would have similarly influenced all conditions. Second, we attempted to control for motor responses associated with hand movements by counterbalancing the hand participants used to advance from word to word. Specifically, half of the participants moved through sentences using their right hand and the other half advanced using their left hand. Finally, we ensured that

⁸This has been interpreted as reflecting a difficulty of “wrapping-up” sentences with anomalies or as related to decision making.

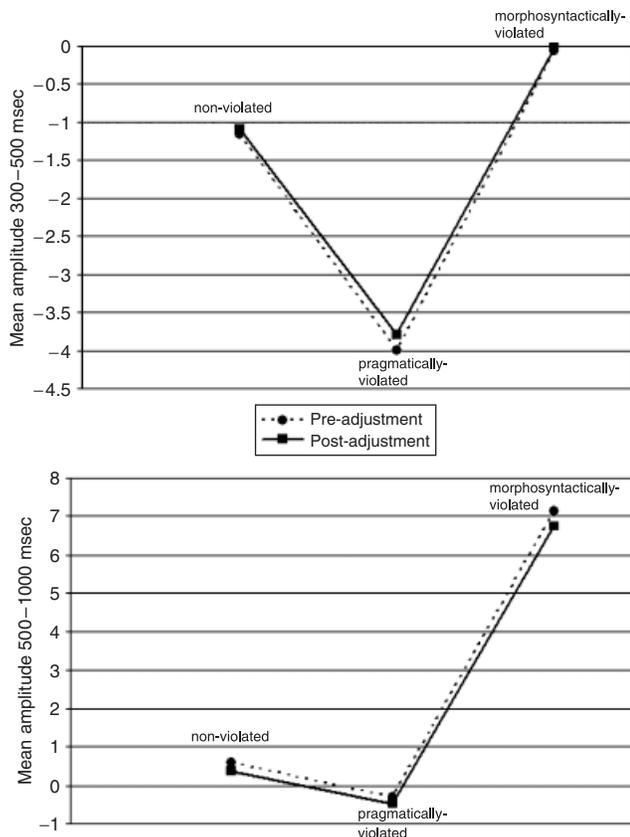


Figure 5. Mean amplitude evoked to the three sentence types at the critical word in the N400 time window (300–500 ms) and the P600 time window (500–1000 ms) before and after adjusting for potential neural contamination due to early component latency differences.

participants sat very still during the experiment. Prior to the beginning of the experiment as well as at each break, participants were instructed to only move their thumbs to advance to the next word and to be sure not to move any other part of their body. This motor response required a very slight movement to activate a small game-pad button.

The results of the current experiment underscore the necessity of employing multiple techniques, either in separate experiments or concurrently as in the present study, in order to fully evaluate the processes involved in language comprehension. As discussed in the Introduction, recent studies have demonstrated that word presentation rate determines the relative influences of lexico-semantic and higher level sentential context on the N400 amplitude, with fast presentation rates associated with greater emphasis on lexico-semantic associations and slow presentation rates linked to influences of higher-level context (e.g., Camblin et al., 2007; Ledoux et al., in press; Svaab et al., 2004). Thus, allowing participants to self-pace through a sentence while ERPs are recorded allows for the examination of the neural correlates of language comprehension at each reader's natural pace, providing increased validity that ERP effects observed reflect neurocognitive processes occurring during normal reading comprehension.

There are additional advantages to using this combined methodology. First, it ensures that participants experience the same conditions during reading time and ERP data collection. It thus allows for direct comparison across the two data sets in the

same individuals. This contrasts with the comparison of data from reading time and ERP data *across* experiments, which requires the assumption that participants in both experiments have similar experiences. Specifically, ERP participants endure a 30–45-min setup prior to beginning an experiment, these participants are seated in a dimly lit room, and they cannot blink during sentence presentation. Thus, differences across experiments can sometimes reflect differences in task demands rather than differences in actual language comprehension processes.

Second, by having two sets of data from the same time point, one can examine correlations between reading time and ERP measures. This allows for direct comparisons between the techniques themselves. Specifically, what are the differences in underlying neurocognitive processes that are being indexed by reading times versus ERP measures? The present reading time/ERP data highlight the differences between cognitive processes indexed by the two techniques. Whereas the ERP data demonstrate that readers immediately (i.e., on the violated critical word) detect both morphosyntactic and pragmatic violations, reading time data suggest that such violations introduce processing load at different points during sentence comprehension—at the anomaly itself for the morphosyntactic violations and at the sentence-final word for the pragmatic violations. Interestingly, both N400 and P600 amplitudes evoked to critical words predicted reading times to sentence-final words. Specifically, larger N400 amplitude predicted longer sentence-final word reading times whereas larger P600 amplitude predicted shorter sentence-final word reading times.

Third, importantly, in a debriefing session, all participants who had previously participated in other non-self-paced ERP studies reported feeling more comfortable reading sentences at their own pace rather than have words “flash at them.” Thus, as intended, the present technique allowed participants to read at a more comfortable pace.

As with all methodologies, there are inherent limitations. People read at different rates, with some individuals reading as rapidly as 250 ms/word and other participants requiring twice that amount of time. In the present study, the average reading times of participants differed by up to 460 ms. Additionally, even within the same participants, reading time differences exist, with most participants reading faster toward the end of the experimental session. There may be several consequences of such individual reading time variability on the ERP components evoked. One is the absence of neural indices of words following the critical word (if the critical word is midsentence). This is apparent by an examination of Figure 3, in which early components to words following the critical word are absent. Thus, in order to examine subsequent words, one must time-lock to their presentation. A second consequence is that reading time differences between conditions could potentially confound the amplitude of ERP effects due to differences in the latencies of early components to subsequent words. Thus this approach may make examining effects that may extend beyond one word more difficult. It is important to note that differences in the latencies of early components to subsequent words do not appear to be confounding ERPs to the critical word in the present study in which, as mentioned above, large individual differences in reading times were observed. First, in overall analyses we replicated previous findings of a larger N400 amplitude to semantically/pragmatically violated critical words and a larger P600 amplitude to morphosyntactic violations relative to critical words in nonviolated sentences (e.g., De Vincenzi et al., 2003; Kuperberg et al., 2003,

2006). Second, even when subtracting out differences in early components taking into account latency differences to subsequent words, we replicated the N400 and P600 effects. However, it is still possible that even though the exogenous components were not observed in the waveforms, they still impacted differences between conditions.⁹ It is imperative, therefore that future experiments examine the extent to which the displacement of visual offset of the critical word as well as visual onset of the following word due to average reading time differences can result in spurious differences in the ERPs.

Finally, although more natural than traditional RSVP displays, this word-by-word self-paced reading approach is still removed from normal reading, with words appearing one by one in the center of the screen. Thus, the next stage is to see if ERPs can be combined with even more naturalistic self-paced reading par-

adigms, such as moving windows, in which words appear from left to right on the screen (Haberlandt, 1994). Horizontal eye movements may, however, preclude this technique from being useful when combined with ERPs.

The present experiment opens the door to future directions in language research. This dual methodology is sensitive to both quantitative and qualitative differences during online language comprehension with millisecond resolution while providing a more natural reading environment for participants. By doing so, it provides increased validity that observed processes are occurring during normal reading comprehension. We believe the combination of self-paced reading and ERP methodologies will bring us closer to a more complete understanding of the processes involved in language comprehension as well as bridging a connection between different methodologies.

REFERENCES

- Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146*, 172–184.
- De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language, 85*, 280–296.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.
- Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.
- Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*, 126–137.
- Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook in psycholinguistics* (pp. 1–31). San Diego, CA: Academic Press, Inc.
- Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114*, 357–374.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30*, 47–61.
- Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. (2006). Neural correlates of processing semantic, syntactic, and thematic relationships. *Language and Cognitive Processes, 21*, 489–530.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 217*, 117–129.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic anomalies. *Nature, 307*, 161–163.
- Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook in psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press, Inc.
- Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (in press). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory and Cognition*.
- Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading—XVI. The visual span in normal and low vision. *Vision Research, 37*, 1999–2010.
- Marslen-Wilson, W. D., Brown, C., & Tyler, L. K. (1988). Lexical representations in language comprehension. *Language and Cognitive Processes, 3*, 1–17.
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19*, 668–682.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118*, 155–167.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.
- Osterhout, L., & Holcomb, P. J. (1995). Event-related potentials and language comprehension. In M. Rugg & M. Coles (Eds.), *Electrophysiology of mind: Event related brain potentials and cognition* (pp. 171–215). New York: Oxford University Press.
- Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 283–317.
- Rugg, M., & Coles, M. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press.
- Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience, 16*, 715–726.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285–318.

(RECEIVED March 5, 2007; ACCEPTED June 22, 2007)

⁹We thank Jos van Berkum for helpful discussions concerning this potential problem concerning visual components.