Mem Cogn (2014) 42:275–291
DOI 10.3758/s13421-013-0356-z

Distinguishing lexical- versus discourse-level processing using event-related potentials

Yi Ting Huang · Joseph Hopfinger · Peter C. Gordon

Published online: 12 October 2013

Abstract Two experiments examine the links between neural patterns in EEG (e.g., N400s, P600s) and their corresponding cognitive processes (e.g., lexical access, discourse integration) by varying the lexical and syntactic contexts of co-referential expressions. Experiment 1 examined coreferring expressions when they occurred within the same clause as their antecedents (*John/Bill warmly dressed John*). Experiment 2 examined between-clause co-referencing with expressions that also varied in lexical frequency (*John/Weston went to the store so that John/Weston could buy milk*). Evidence of facilitated lexical processing occurred after repeated names, which elicited smaller N400s, as compared with new names. N400s were also attenuated to a greater degree for low-frequency expressions than for high-frequency ones. Repeated names also triggered evidence of postlexical processing, but this emerged as larger P600s for within-clause co-referencing and delayed N400s for between-clause co-referencing. Together, these results suggest that linguistic processes can be distinguished through distinct ERP components or distinct temporal patterns.

Keywords Event-related potentials · Neurolinguistics · Co-reference · N400s · P600s

Introduction

Language comprehension is often characterized as rapid, incremental, and opportunistic—exploiting multiple cues from various sources to resolve ambiguity and make predictions about upcoming material (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Yet there remains debate about how these cues are processed during real-time interpretation. One class of hypotheses argues that inputs are mapped onto separable levels of representations and that processes at preceding levels partially constrain computations at subsequent levels (Ferreira & Patson, 2007; Huang & Gordon, 2011; Tily, Federenko, & Gibson, 2010). These accounts maintain that architectural features of the cognitive system limit the concurrent processing of all information, resulting in differential time courses for different types of linguistic procedures. For example, since lexical processes logically precede discourse processes, one would expect evidence of the former to emerge prior to the latter (Huang & Snedeker, 2009a, 2009b, 2011; Ledoux, Gordon, Camblin, & Swaab, 2007). A contrasting class of hypotheses suggests that all relevant inputs are recruited by an all-purpose parser that constrains comprehension in an optimal and exhaustive manner (Grodner, Klein, Carbary, & Tanenhaus, 2010; Jurafsky, 1996; Levy, 2008). These accounts contend that any available information that *can* inform sentence interpretation *will* be used to affect processing in its *earliest* moments. The type of massive interactivity suggests that time course information is an unreliable indicator of underlying linguistic processes.

Much of the work motivating these theories has come from behavioral measures. Thus, the challenge remains for the field to develop a cognitive theory of language comprehension that is also consistent with evidence from neural research. Neural measures, such as event-related potentials (ERPs), have provided valuable assessments of language comprehension precisely because they provide fine-grained time course information that is sensitive to different language processes. For example, one prominent component has been the N400, a negative polarity deflection in the ERP waveform that peaks

Y. T. Huang (✉)
Department of Hearing and Speech Sciences, University of Maryland College Park, 0100 Lefrak Hall, College Park, MD 20742, USA
e-mail: ythuang1@umd.edu

J. Hopfinger · P. C. Gordon
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

approximately 400 ms after stimulus onset. Since this component is triggered by factors affecting word recognition, such as frequency (Allen, Badecker, & Osterhout, 2003; Van Petten & Kutas, 1990), repetition (Ledoux et al., 2007; Rugg & Nagy, 1987), predictability (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984), and semantic relatedness (Kutas & Hillyard, 1980; Rugg, 1985), it has traditionally been thought to index lexical-level processes. In contrast, the P600 is a positive-going component that peaks approximately 600 ms after stimulus onset. This component is triggered by mismatches in word order (Hagoort, Brown, & Groothusen, 1993; Osterhout, 1997) and gender/number morphology (Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995), and thus it has traditionally been linked to syntactic processes.

Despite the appealing clarity of this division between ERP components, recent studies have uncovered notable patterns that challenge this straightforward interpretation. For example, Swaab, Camblin, and Gordon (2004; Ledoux et al., 2007) recorded ERPs to sentences like (1). Here, critical words were either new names (*John* following *Bill*) or repeated names (*John* following *John*) that were preceded by either prominent (*John*) or nonprominent antecedents (*John and Neil*).

(1) a. PROMINENT/REPEATED: John left work after **John** completed the project
b. PROMINENT/NEW: Bill left work after **John** completed the project
c. NONPROMINENT/REPEATED: John and Neil left work after **John** completed the project
d. NONPROMINENT/NEW: Bill and Neil left work after **John** completed the project

Gordon, Swaab, and colleagues found that following nonprominent antecedents, repeated names yielded reduced N400s, as compared with new names (1c vs. 1d) (Ledoux et al., 2007; Swaab et al., 2004). This repetition priming demonstrates that prior exposure to a word facilitates subsequent processing of the same word (Liversedge et al., 2003; Raney, Therriault, & Minkoff, 2000; Rugg & Nagy, 1987; Traxler, Foss, Seely, Kaup, & Morris, 2000) and is consistent with traditional interpretations of the N400 as an index of lexical access (Brown & Hagoort, 1993; Chwilla, Brown, & Hagoort, 1995; Holcomb, 1993; Lau, Phillips, & Poeppel, 2008; Rugg, Furda, & Lorist, 1988; Van Berkum, Hagoort, & Brown, 1999; Van Petten & Kutas, 1991;). However, in the same studies, repeated names also generated greater N400s when they co-referenced prominent antecedents, relative to nonprominent ones (1a vs. 1c). This *repeated-name penalty* demonstrates that the use of an overinformative, repeated expression interferes with discourse integration (Almor, 1999; Gordon & Hendrick, 1998).

Altogether, these findings are notable for two reasons. First, the same repeated expression generated N400 responses at both the lexical and discourse levels, with no apparent difference in the time course of the processes that generated the effects. This is at odds with behavioral evidence from reading-time studies, which demonstrated that discourse processes indexed by the repeated-name penalty occur well after the word recognition processes indexed by repetition priming (Ledoux et al., 2007). Second, when lexical and discourse effects were compared head-to-head in the prominent sentences, there were no clear differences in the ERP components generated by repeated and new names (1a vs. 1b). The absence of an N400 difference in this comparison raises questions about the exact relationship between the lexical and discourse processes in comprehension. Are linguistic inputs processed via distinct levels of representations (e.g., lexical, discourse), or are they analyzed via a single level of representation (e.g., meaning)?

Similar puzzles have emerged in studies on the interface between lexical semantics and syntax. Kuperberg, Sitnikova, Caplan, and Holcomb (2003) examined sentences like (2), which varied the semantic relatedness (related vs. unrelated) and presence of a thematic-role violation (no violation vs. violation) expressed in the relationship between preceding subject nouns (*boys*/*eggs*) and subsequent verbs (*eat/plant*).

(2) a. RELATED/NO VIOLATION: For breakfast the boys would **eat** toast and jam
b. UNRELATED/NO VIOLATION: For breakfast the boys would **plant** flowers in the garden
c. RELATED/VIOLATION: For breakfast the eggs would **eat** toast and jam
d. UNRELATED/VIOLATION: For breakfast the eggs would **plant** flowers in the garden

They found a greater N400 for unrelated, as compared with related, verbs when there was no thematic role violation (2b vs. 2a) (Kim & Osterhout, 2005; Kuperberg et al., 2003; Kuperberg et al., 2007). This semantic relatedness effect is consistent with accounts of the N400 as reflecting either lexical access (Kutas & Hillyard, 1980, 1984; Rugg, 1985) or postlexical integration (Brown & Hagoort, 1993; Ledoux et al., 2007; Swaab, Camblin, & Gordon, 2004). However, Kuperberg et al. (2007) also found that thematic role violations caused a P600 effect but no N400 differences regardless of whether the subject noun was semantically related to the verb (2c and 2d vs. 2a). These *semantic P600s* are surprising for two reasons. First, none of the sentences in (2) violated the syntactic dimensions typically associated with P600s—for example, nonconventional word order, agreement errors, or morphological mismatches (Hagoort et al., 1993; Osterhout & Holcomb, 1992). Second, the absence of a larger N400 response in sentences featuring both unrelated words *and* thematic role violations is puzzling, since prior research has shown evidence of both components within a single sentence (Osterhout & Nichols, 1999). Both patterns appear to be at odds with work showing that N400s and P600s are moderated by different aspects of language. One possibility is that these effects provide evidence in favor of a massively interactive comprehension system. For example, Kim and Osterhout (2005) contended that the semantic P600s demonstrate that semantic and syntactic interpretations are processed in parallel and that the former can influence the latter when they are sufficiently robust. Similarly, Kuperberg and colleagues (2007) suggested that the failure to find *both* N400s and P600s in the unrelated/violation sentences reflects the canceling of semantic integration in the presence of syntactic violations. Accounts such as these have been prominent in the neurolinguistics literature (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Nieuwland & Van Berkum, 2006a, 2006b; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003).

To summarize, previous studies have suggested clear divisions between neural patterns (e.g., N400s, P600s) and their corresponding cognitive processes (e.g., lexical processing, postlexical integration, syntactic processing). However, recent work from two different linguistic domains raises questions about these traditional mappings (Kim & Osterhout, 2005; Kuperberg et al., 2003; Kuperberg et al., 2007; Ledoux et al., 2007; Swaab et al., 2004). Thus, it remains unclear how the comprehension system can generate both distinct neural and temporal patterns for some linguistic phenomena but conflated patterns for others. One possibility is that these recent results provide definitive evidence in favor of a single level of processing. This type of massive interactivity across all relevant inputs suggests no principled distinction between lexical and postlexical effects. However, a second possibility is that levels of language processing function in separable ways, and interactions are limited to operations at the interface of related linguistic representations. For example, the use of proper names by Gordon, Swaab, and colleagues may offer an exceptional case where referring expressions are uniquely linked to discourse representations (Kripke, 1980). Similarly, the study of thematic role assignments by Kuperberg and colleagues highlight a case where animacy cues from lexical semantics are highly correlated with syntactic categories (Jackendoff, 1972; Ladusaw & Dowty, 1988).

To distinguish these possibilities, the present study examines whether highly correlated lexical and discourse processes can *still* be distinguished during comprehension, through the manipulation of syntactic and lexical factors. Experiment 1 includes a critical new manipulation of syntactic context, contrasting critical expressions that occur within a clause (e.g., *John warmly dressed John*) with those that occur between clauses (e.g., *John left work after John completed the project*). Experiment 2 focuses on a lexical factor, contrasting high-frequency names (e.g., *John*) with low-frequency ones (e.g., *Earl*). The goal of both experiments is to assess the neural and temporal patterns associated with discourse-level interpretations of co-referential expressions by comparing them with the reduced N400 response following repeated names, as compared with new names (Ledoux et al., 2007; Swaab et al., 2004). Repetition priming of this kind provides a useful benchmark of lexical processing with which the timing of discourse processing can be compared (Huang & Gordon, 2011). If linguistic inputs are analyzed via separable levels of interpretation, lexical and postlexical processes should correspond to distinct neural and/or temporal patterns. However, if inputs are analyzed via a single level of interpretation, overlapping patterns should continue to persist.

Experiment 1

Experiment 1 examined ERPs to a target referential expression (e.g., *John* in 3a and 3d) as a function of an earlier referential expression, which we will call *Noun1*. Noun1 could be a simple noun phrase (NP), in which case its referent was prominent (e.g., 3a and 3b), or it could be embedded as the possessor within a complex possessive NP, in which case its referent was nonprominent (e.g., 3c and 3d). This characterization of the relationship between referential prominence and syntactic structure follows the analysis of Gordon and Hendrick (1998) in which referential prominence was defined as inversely related to syntactic embedding. In addition, Noun1 could be the same as the target (the *repeated* condition; e.g., 3a and 3c), or it could differ from the target (the *new* condition; e.g., 3b and 3d).

(3) a. PROMINENT/REPEATED: Yesterday John warmly dressed **John** before school
 b. PROMINENT/NEW: Yesterday Bill warmly dressed **John** before school
 c. NONPROMINENT/REPEATED: Yesterday John's mother warmly dressed **John** before …
 d. NONPROMINENT/NEW: Yesterday Bill's mother warmly dressed **John** before school

Note that while a repeated expression can felicitously co-refer with a nonprominent antecedent (e.g., 3c), it cannot with a prominent one (e.g., 3a). Instead, a reflexive pronoun (*himself*) is required for co-reference (or anaphora) in such cases where the two expressions are in the same clause (Chomsky, 1981; Gordon & Hendrick, 1997).

As was discussed above, previous studies have shown that repeated expressions that affect comprehension at both the lexical and discourse levels generate N400 responses that are indistinguishable (Ledoux et al., 2007; Swaab et al., 2004). Contrary to previous behavioral research, this suggests that there are no differences in the neural or temporal patterns associated with language processing at these two different levels. However, another possibility is that lexical and discourse processes are, in fact, distinct components of comprehension but evidence of these separate generators is obscured in situations where there is complete overlap in the ERP components. However, recent work suggests that it may be possible to eliminate this overlap through a manipulation of syntactic context. Gordon, Kacinik, and Swaab (2013) found that when repetition occurred *within* a single clause, a prominent antecedent (3a) generated a greater P600, as compared with a nonprominent one (3c). This effect is consistent with what has been found following reflexives that do not match their prominent antecedents (Osterhout & Mobley, 1995; e.g., *herself* rather than *himself* following a stereotypically male name).

Critically, the presence of a P600 (instead of an N400) during discourse processing allows us to distinguish between different accounts of the relationship between lower-level lexical processes and higher-level discourse effects. Since the interpretation of repeated expressions triggers distinct components across different processes, a direct comparison of their neural responses within a sentence will assess whether lexical effects can co-occur with discourse effects or whether the presence of one necessarily cancels the other (Kuperberg et al., 2003; Kuperberg et al., 2007; Ledoux et al., 2007). Thus, unlike Gordon and colleagues (2013), the present study also manipulates whether prominent antecedents are co-referenced through repeated or new names (3a vs. 3b). If lexical and discourse processes are separable during comprehension, we would expect evidence of both N400 and P600 responses following a repeated expression, as compared with a new one. However, if these processes are largely overlapping, the presence of a syntactic violation may lead to a top-down cancelation of lexical processes (Kim & Osterhout, 2005; Kuperberg et al., 2007). If this were the case, we would find evidence of a P600 response, but not an N400 response.

Method

Participants

Sixteen right-handed adults participated in this study. They were recruited from the university population at the University of North Carolina at Chapel Hill and were compensated $25 for their participation. All participants were native English speakers, and none had any history of neurological impairment. Written consent was obtained from each participant prior to beginning the study.

Procedure

Participants sat in an arm chair inside a dimly lit room that was electrically shielded and sound attenuated. A computer screen was placed in front of the participants, approximately 55 cm away from their eyes. An eyetracker was placed below the screen and was used to monitor participants' blinking and eye movements throughout the study. At the beginning of the study, the experimenter told participants that they would be asked to judge whether the sentence sounded "good" or "bad." They were told to base these judgments on their intuitions of how they imagined most people would speak, rather than on any prescribed notions of what is proper or correct. On each trial, the words for the sentence would appear one at a time. The experimenter emphasized that it was important for participants to refrain from moving or blinking during the presentation of the sentence. Responses to the grammaticality judgments could be made by pressing one of two buttons on a video game consol. After their response, participants were given the opportunity to blink and rest their eyes. When they were ready to proceed, they could press any button to continue onto the next trial.

Each trial began with a fixation cross that appeared at the center of the screen for 1,000 ms. This cross alerted participants to the beginning of the trial and also marked the location of the subsequent words of the sentence. These words appeared in rapid serial visual presentation (RSVP) with an on-screen duration of 300 ms per word and an interstimulus interval of 200 ms. They were presented against a black background in 70-point white Tahoma font. Unlike the words for the sentences, the words for the judgment question ("Good or bad?") appeared on the screen simultaneously, after the completion of the sentence, and remained there until a response was made.

Materials

The materials for the four critical trial types follow the example in sentence (3) and represent the cells of a 2 × 2 design. The first factor, *prominence*, contrasts the use of a prominent, singular NP (*John*) versus a nonprominent, possessive NP (*John's brother*) as antecedents. The second factor, *repetition*, contrasts the repetition of a previously mentioned name (*John . . . John*) versus the introduction of a new one (*Bill . . . John*) as co-referring expressions. Both factors were varied within subjects.

The sentence frames were adopted from Gordon et al. (2013). These frames included locative phrases at the beginning and verb phrases at the end, to ensure that names would not appear in sentence-initial or sentence-final positions. Both first and second mentions of names appeared in the same clause. This created a situation where co-reference of an antecedent was most felicitously done through a reflexive pronoun (e.g., *himself*, *herself*). Sentences varied in length from 8 to 11 words, with a mean length of 9.9 (SD = 0.8). Examples of the critical stimuli are presented in Appendix 1. Four versions of each critical base item were used to create four presentation lists, such that each list contained 60 items in each condition and each base item appeared just once in every list.

These 240 critical trials were randomized with 80 control trials and 10 practice trials. Control items were of similar character to the critical items but used the reflexive pronoun in place of a critical name. On half the trials, congruent sentences were ones where the gender of the reflexive matched the gender of its antecedent, both prominent and nonprominent (*John . . . himself*, *John's sister . . . herself*). On the other half, incongruent sentences were ones where the gender of the reflexive did not match its antecedent (*John . . . herself*, *John's sister. . . himself*). Since prior work has shown that P600s are elicited by these kinds of gender mismatches between antecedents and pronouns (Osterhout & Mobley, 1995), these control trials offer an informative benchmark for effects that may emerge in the critical trials. Practice items varied the presence or absence of morphological errors (e.g., verb tense, number agreement). A total of 330 trials were divided into one practice block and eight test blocks. Each test block lasted about 10 min.

EEG recordings

The EEG was recorded from 96 electrodes (see Fig. 1 for layout) fitted into an elastic cap using an ActiveTwo EEG system with active electrodes (BioSemi; Amsterdam). Activity associated with eye movements was monitored through four additional electrodes at the suborbital region and outer canthi of the right and left eyes. Online recordings were single-ended potential measurements with respect to a common mode sense site near the vertex, and data were sampled at 256 Hz. The data were referenced, offline, to the average of the activity recorded from the left and right mastoids, and the signal was filtered with a bandpass of 0.01–30 Hz. EEGs were analyzed using the Brain Electrical Source Analysis (BESA) 3.0 software package. Initial processing screened single-trial waveforms for artifacts such as amplifier blocking and muscle and eye movements over an epoch beginning from −100 ms before the critical word to 1,000 ms after the critical word. Trial rejection rates on the basis of artifacts for individual subjects varied from 6 % to 22 % of critical and control trials, with an average of 11 % across subjects. ERPs for each participant were calculated by averaging over artifact-free trials for the critical word in the critical and control conditions.

Results

Behavioral data

Accuracy of the grammaticality judgments ranged from 68 % to 93 % across subjects, with the mean performance at 83 % (*SD* = 8 %). There were no significant differences in performance across conditions, Fs < 1.00, ps > .50. This confirms that participants were able to correctly distinguish grammatical sentences (e.g., 3b–d) from ungrammatical ones (e.g., 3a).

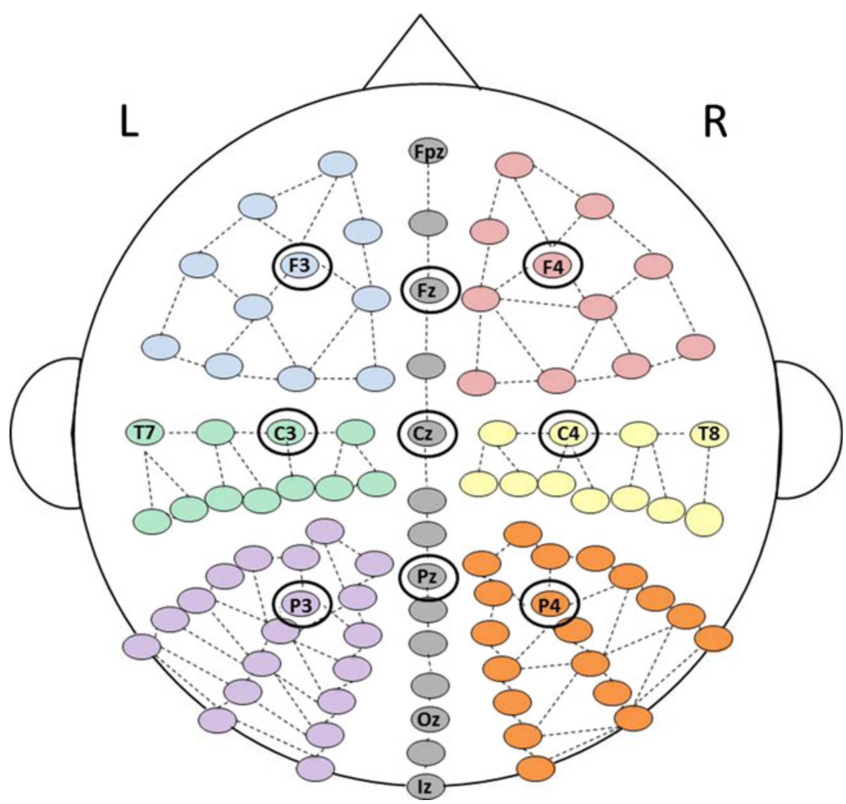

Fig. 1 Electrode montage from a 96-channel cap. All electrodes were analyzed on the basis of seven areas of interest: midline (gray), left-frontal (blue), right-front (red), left-central (green), right-central (yellow), left posterior (purple), and right-posterior (orange). Electrodes corresponding to illustrated waveforms are indicated with dotted circles

ERP analysis

The mean amplitudes of the ERPs to the critical words were analyzed using a series of repeated measures analyses of variance (ANOVAs) during the N400 (250–500 ms) and P600 (650–800 ms) time windows, time-locked to the onset of the target name in the critical trials and to the onset of the reflexive pronoun in the control trials. The latency ranges were based on visual inspection of the waveforms within the time windows reported previously for the N400 and P600 components. In the critical trials, omnibus analyses were first conducted over three within-subjects variables: prominence (prominent vs. nonprominent), repetition (new vs. repeated), and electrode region (seven regions). The latter variable divided electrode sites on the basis of their hemisphere (left vs. right) and anteriority (frontal vs. central vs. posterior), resulting in six regions (left-frontal, left-central, left-posterior, right-frontal, right-central, right-posterior) plus the midline region. In the control trials, omnibus analyses were conducted over three within-subjects variables: prominence (prominent vs. nonprominent), congruency (incongruent vs. congruent), and electrode region. Significant interactions between the manipulated variables and region were followed up with planned comparisons, focusing on the corresponding electrodes. Greenhouse–Geisser corrections were applied to compensate for inhomogeneous variance and covariance across treatment levels (Greenhouse & Geisser, 1959). Adjusted p-values are reported below.

N400 The omnibus ANOVA of the N400 following the critical name revealed a significant two-way interaction between repetition and electrode region, $F(6, 90) = 3.60$, $p < .05$, but no additional three-way interaction between repetition, prominence, and electrode region ($p > .15$). Planned comparisons revealed that repetition effects were maximal over midline electrodes, where new names elicited greater negativity, as compared with repeated names, $F(1, 15) = 19.40$, $p < .01$. Figures 2 and 3a illustrate that this repetition priming was observed in both nonprominent trials, where the use of the repeated name was felicitous, $F(1, 15) = 14.72$, $p < .01$, and prominent trials, where it was infelicitous, $F(1, 15) = 9.31$, $p < .01$. There was no additional main effect of or interaction with prominence (both ps > .30).

P600 The omnibus ANOVA of the P600 following the critical name revealed a significant three-way interaction between

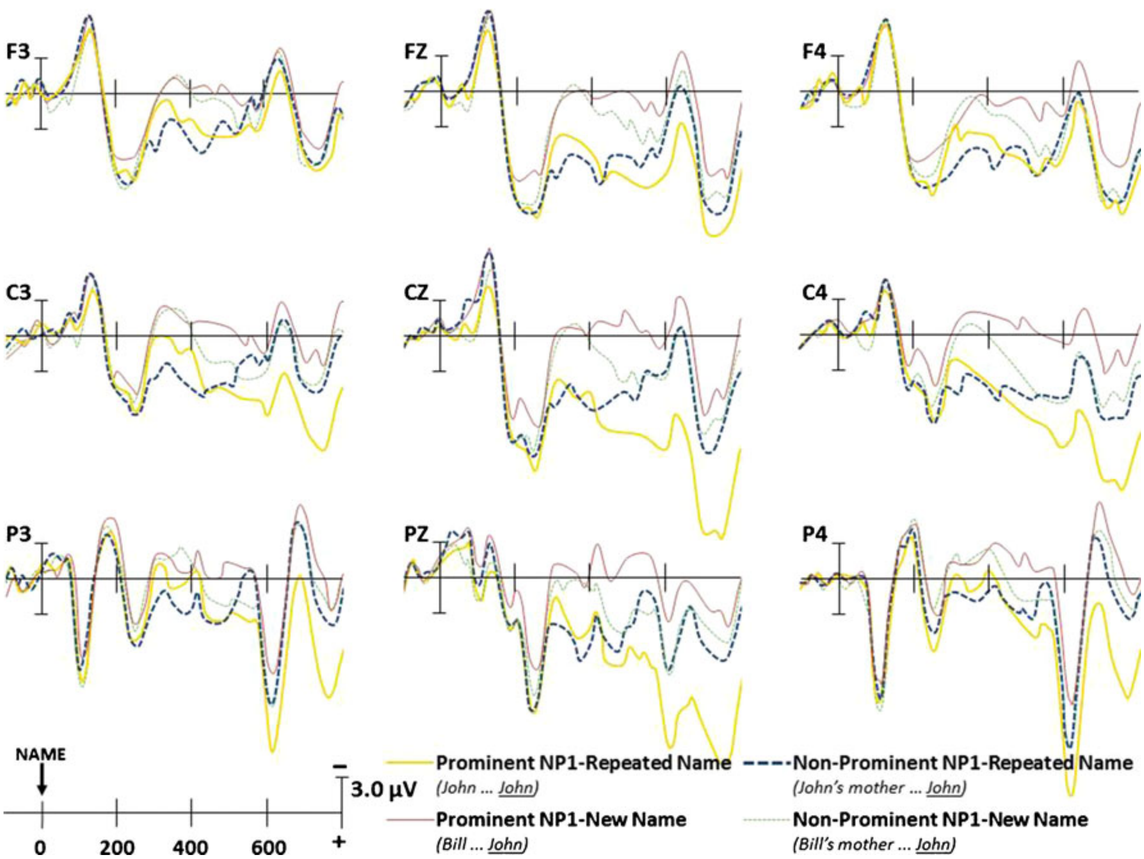

Fig. 2 In Experiment 1, the effects of antecedent type (prominent vs. nonprominent) on the interpretation of co-referring expressions (repeated vs. new names). The ERPs were time-locked to the critical name and reflect grand averages across all participants, recorded from frontal (F3, Fz, F4), central (C3, Cz, C4), and posterior (P3, Pz, P4) sites

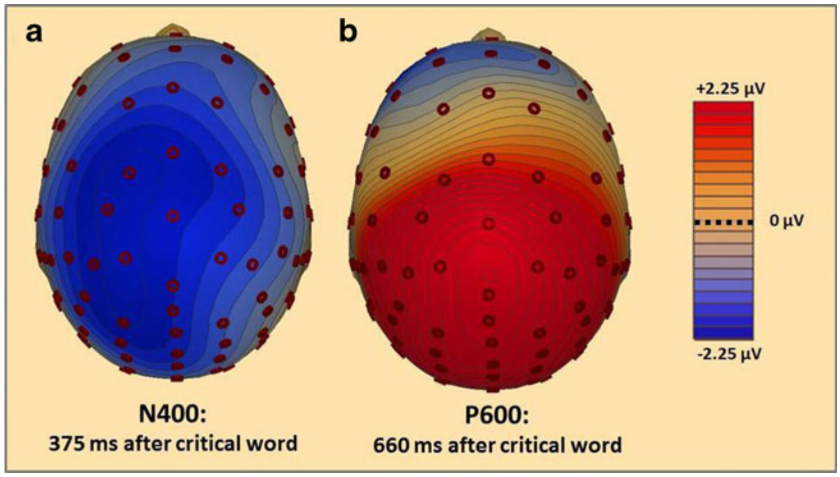

Fig. 3 In Experiment 1, the voltage maps illustrate the spatial distribution of the ERP responses. On critical trials, the **a** N400 response 375 ms after the critical name reflects the new names minus repeated names difference in the nonprominent condition, and the **b** P600 response 660 ms after the critical name reflects the prominent names minus nonprominent names difference in the repeated condition

repetition, prominence, and electrode region, $F(6, 90) = 5.31$, $p < .05$. Planned comparisons revealed that interactions between repetition and prominence were maximal over right-central electrodes, $F(1, 15) = 10.45$, $p < .01$. Figures 2 and 3b illustrate that while new and repeated names were equivalent in the nonprominent condition, $F(1, 15) = 1.05$, $p > .30$, repeated names generated greater positivity than did new names in the prominent condition, $F(1, 15) = 16.10$, $p < .01$. This pattern also led to a main effect of repetition, $F(1, 15) = 16.19$, $p < .01$. There was no additional main effect of prominence ($p > .15$).

Finally, the P600 effect on the critical trials resembled the P600 generated on control trials, where a gender mismatch between the antecedent and the reflexive pronoun produced a standard P600 response of the sort observed by Osterhout and Mobley (1995). The omnibus ANOVA revealed a significant two-way interaction between congruency and electrode region, $F(6, 90) = 7.66$, $p < .01$, but no additional three-way interaction between congruency, prominence, and region ($p > .60$). Planned comparisons revealed that incongruency effects were maximal over the left posterior electrodes, where incongruent pronouns elicited greater positivity, as compared with congruent pronouns, $F(1, 15) = 24.51$, $p < .001$. This occurred following both prominent antecedents, $F(1, 15) = 12.31$, $p < .01$, and nonprominent antecedents, $F(1, 15) = 14.91$, $p < .01$. There was no additional main effect of or interaction with prominence (both ps > .30).

Discussion

In Experiment 1, we distinguished between lexical processes and discourse processes by varying the prominence and repetition of names within a clause and measuring their effects on two neural responses, the N400 and the P600. Consistent with prior research, we found that new names were more difficult to process, generating larger N400s, as compared with repeated names (for other types of words, see Ledoux et al., 2007; Rugg & Nagy, 1987). This repetition priming demonstrates that prior recognition of a word facilitates subsequent processing of the same word. Critically, we also found evidence that the infelicitous co-reference created by using a repeated expression to refer to a prominent antecedent within the same clause resulted in a P600. This P600 effect was similar to that found in cases of gender mismatches between pronouns and their antecedents (Gordon et al., 2013; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995). Thus, the repeated name in sentence 2a not only produced a facilitative effect on the N400, as compared with new names, but also produced an inhibitory effect on the P600, as compared with the other sentence types.

Critically, the finding that the same repeated expression leads to both patterns within the same sentence demonstrates a novel finding of *distinct* effects on lexical *and* discourse processing. This sheds light on how to interpret prior evidence of interactivity. In particular, it provides new evidence suggesting that while lexical and postlexical processes may be highly correlated at linguistic interfaces (Kuperberg et al., 2003; Kuperberg et al., 2007; Ledoux et al., 2007), they can also function independently for other aspects of interpretation. As the present findings show, when lexical and discourse effects are associated with different components (N400 and P600, rather than N400 for both), evidence for both processes is revealed. Similarly, when lexical and postlexical effects are associated with different cognitive procedures (lexical access and co-reference, rather than thematic role assignment for both), evidence of two processes is revealed. We will return to the implications of these patterns on models of language comprehension in the General Discussion section.

Experiment 2

In Experiment 2, we wanted to examine more closely an important question raised by these results: What is the cause of the N400? In particular, while the P600s elicited by repeated mention within a clause suggest that neural responses are sensitive to structural features of co-reference (Gordon et al., 2013), there remains the puzzle of why N400s index both lexical and discourse effects when co-reference occurs between clauses (Ledoux et al., 2007; Swaab et al., 2004). Recall that in the original studies, repeated names that co-referenced nonprominent antecedents generated reduced N400 responses, as compared with new names (repetition priming; compare 1c to 1d). However, repeated names also generated greater N400 responses when they co-referenced prominent antecedents, relative to nonprominent antecedents (repeated name penalty; compare 1a to 1c). Yet these two cases naturally reflect very different procedures. This intuition is confirmed by behavioral studies demonstrating evidence of rapid repetition priming (Liversedge et al., 2003; Raney et al., 2000; Traxler et al., 2000) paired with delayed effects of repeated name penalty (Ledoux et al., 2007; Swaab et al., 2004).

Experiment 2 distinguishes between the neural processes associated with the N400 by examining possible interactions with lexical frequency. Like (1), critical sentences vary the prominence of antecedents and the repetition of referring expressions between two clauses of a sentence. However, unlike (1), they also recruit names that are either low or high frequency [see (4) and (5), respectively].

(4) a. LOW/PROMINENT/REPEATED: Yesterday Earl left work after **Earl** completed the project
  b. LOW/PROMINENT/NEW: Yesterday Wade left work after **Earl** completed the project
  c. LOW/NONPROMINENT/REPEATED: Yesterday Earl and Neil left work after **Earl** . . .
  d. LOW/NONPROMINENT/NEW: Yesterday Wade and Neil left work after **Earl** . . .

(5) a. HIGH/PROMINENT/REPEATED: Yesterday John left work after **John** completed the project
  b. HIGH/PROMINENT/NEW: Yesterday Bill left work after **John** completed the project
  c. HIGH/NONPROMINENT/REPEATED: Yesterday John and Neil left work after **John** . . ..
  d. HIGH/NONPROMINENT/NEW: Yesterday Bill and Neil left work after **John** . . .

At the lexical level, previous behavioral research has found that repetition priming interacts with lexical frequency: Priming is larger for words that are low in frequency and smaller for words that are high in frequency (Lowder, Choi, & Gordon, 2013; Scarborough, Cortese, & Scarborough, 1977; Young & Rugg, 1992). This pattern suggests that repeated mention facilitates word recognition the most in cases where episodic memory representations are least robust. Lexical frequency also interacts with subsequent integration but does so along a different time course, as compared with word recognition (Johnson, Lowder, & Gordon, 2011; Staub, 2011; Tily et al., 2010). For example, Tily and colleagues varied verb frequency in structurally complex object-cleft sentences (*It was Vivian who Terrence lectured/chided for always being late*) and found that high-frequency verbs lead to earlier processing of the syntactic ambiguity on the cleft region (*Terrence lectured*). However, this effect did not emerge in low-frequency verbs until the postcleft region (*for always*). These results suggest that delays in lexical processing have cascaded effects on postlexical integration.

These findings generate predictions for possible effects of lexical frequency on the N400 responses elicited by repetition priming and the repeated name penalty (Ledoux et al., 2007; Swaab et al., 2004). In particular, if the N400 reflects a single neural process, evidence of lexical access and integration should again both be apparent immediately after names across all frequencies. However, if the N400 reflects multiple processes, variations in lexical frequency may distinguish between access and integration. In particular, consistent with the studies above, recent ERP results have demonstrated that difficulty in integration can have downstream effects on neural processing (Hagoort et al., 1993; Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010; Osterhout & Holcomb, 1992). This predicts that delays in lexical processing for low-frequency names may lead to corresponding delays in lexical integration.

Method

Participants

Twenty-four right-handed adults participated in this study. They were recruited from the university population at the University of North Carolina at Chapel Hill and were compensated $25 for their participation. All participants were native English speakers, and none had any history of neurological impairment. Written consent was obtained from each participant prior to beginning the study.

Procedure

The procedure was similar to that in Experiment 1. However, in this experiment, participants were asked true–false comprehension questions after each sentence. Responses were made by pressing one of two buttons on a video game consol.

Materials

The materials for the eight critical trial types follow the example in sentences (4) and (5) and represent the cells of 2 × 2 × 2 design. The first factor, prominence, contrasts the use of a prominent, singular NP (*John*) versus a nonprominent, conjoined NP (*John and Neil*) in the first clause of the sentence. The second factor, repetition, contrasts the use of a critical name that repeats a previously mentioned one (*John . . . John*) versus introduces a new one (*Eric . . . John*) in the second clause of the sentence. Finally, the third factor, frequency, contrasts the use of a low-frequency critical name (*Earl*) versus a high-frequency one (*John*). All three factors were varied within subjects.

The names were selected from a database of names of all first-year undergraduate students who enrolled at the University of North Carolina at Chapel Hill over a 5-year period (total token frequency = 28,676, total type frequency = 2,668). Half of the selected items were female, and half were male. Across all items, selected names varied in frequency from 4 to 623 tokens, with a mean of 3.8 (*SD* = 0.8) in the low condition and 155.5 (*SD* = 131.0) in the high condition. The names in the conjoined NP were moderate in frequency (*M* = 15.2, *SD* = 3.0) to avoid any bias from nonmanipulated variables. Across all items, the names varied in length from 3 to 11 characters. The mean length of names was 6.1 (*SD* = 1.4) and was matched across conditions to avoid potential confounds with other variables.

The names were embedded into sentence frames adopted from Ledoux et al. (2007). As in Experiment 1, these frames included locative phrases at the beginning and verb phrases at the end, to ensure that names would not appear in sentence-initial or sentence-final position. However, unlike in Experiment 1, they also involved two clauses joined by a temporal or causal connective (e.g., *when*, *after*). In the prominent conditions, this positioning created a situation where co-reference of the antecedent was most felicitously done through a pronoun (e.g., *he*, *she*). In the nonprominent conditions, the position of the prime/antecedent NP (*John*), relative to its conjoined NP (*Neil*), was counterbalanced across items. Sentences varied in length from 10 to 22 words, with a mean length of 15.6 (*SD* = 2.2). Additional examples of the critical stimuli are presented in Appendix 2. Eight versions of each critical base item were used to create eight presentation lists such that each list contained 20 items in each condition and each base item appeared just once in every list. These 160 critical trials were randomized with 80 filler trials and 10 practice trials. These items were of similar character to the critical items but only included names of moderate length and frequency and did not systematically vary across the manipulated factors. The 250 trials were divided up into one practice block and eight test blocks. Each test block lasted approximately 10 min.

Ledoux and colleagues (2007) reported results from offline measures of the plausibility and interpretation of sentences where repeated names were used to co-reference prominent antecedents (see 4a and 5a). Consistent with prior studies (Almor, 1999; Garrod, Freudenthal, & Boyle, 1994; Gordon, Grosz, & Gilliom, 1993; Gordon & Hendrick, 1997; Kennison & Gordon, 1997; Yang, Gordon, Hendrick, & Hue, 2003), participants' ratings revealed that these sentences were significantly less felicitous than repetition of nonprominent names or the introduction of a new name following prominent and nonprominent NPs. However, ratings among the latter conditions did not differ from each other, suggesting that both repeated and new names were plausible references in the discourse context. A second study confirmed that despite the unnaturalness of using a repeated name to refer to a prominent NP, participants consistently interpreted this second mention as co-referential with the first. This suggests that participants' awareness of the infelicity did not prevent them from establishing reference to the antecedent.

EEG recordings

EEGs were recorded using an identical procedure as in Experiment 1. Initial processing screened single-trial waveforms for artifacts over an epoch beginning from −100 ms before the prime/antecedent, critical word (*name*), and the word after (*name + 1*) to 600 ms after each of these points. Trial rejection rates on the basis of artifacts for individual subjects varied from 2 % to 22 % of critical and control trials, with an average of 9 % across all participants. ERPs for each participant were calculated by averaging over artifact-free trials in the eight critical conditions.

Results

Behavioral data

Accuracy of the comprehension questions ranged from 70 % to 90 % across subjects, with the mean performance at 81 % (*SD* = 5 %). Participants were more accurate for comprehension questions involving repeated names, as compared with new names, particularly in the nonprominent condition. This led to a main effect of repetition, $F(1, 23) = 22.35$, $p < .001$, as well as an interaction between repetition and prominence, $F(1, 23) = 5.39$, $p < .05$. These patterns directly reflect the ease of recalling the relevant referents in the sentences: When more names are introduced, it becomes more difficult to remember who did what to whom. Similarly, participants were more accurate for comprehension questions involving high-frequency names, as compared with low-frequency names, leading to a main effect of frequency, $F(1, 23) = 5.62$, $p < .05$. This suggests that less common names were more difficult to encode and retrieve in memory, leading to decreased accuracy in subsequent judgments. All other main effects and interactions were not significant (all *F*s < 1.0, all *p*s > .50).

ERP analysis

ERP data were analyzed using procedures similar to those in Experiment 1. The mean amplitudes of the ERPs to the critical words were analyzed using a series of repeated measures ANOVAs during the N400 (250–500 ms) and P600 (650–

800 ms) time windows, time-locked to the onset of the prime/antecedent, critical name, and critical name + 1 regions. An omnibus analysis was first conducted over the four within-subjects variables: prominence (prominent vs. nonprominent), repetition (new vs. repeated), frequency (low vs. high), and electrode region. A Greenhouse–Geisser correction was applied to these analyses, and the adjusted p-values are reported below.

Prime/antecedent In order to establish that the present frequency manipulation would have the expected effect (enhanced N400 to lower frequency words), we first analyzed the ERPs evoked by the prime/antecedent. The omnibus ANOVA of the N400 region following the prime/antecedent revealed significant main effects of electrode region, $F(6, 138) = 4.16$, $p < .01$, as well as significant interactions between electrode region and frequency, $F(6, 138) = 6.73$, $p < .001$, and prominence, $F(6, 138) = 2.82$, $p < .05$, but no additional three-way interaction between frequency, prominence, and electrode region ($p > .90$). Subsequent analyses explored these effects in greater detail by collapsing over levels of repetition, since these differences would not yet be apparent at the onset of the antecedent. Planned comparisons revealed that frequency effects were maximal over left and right posterior electrodes, where low-frequency names elicited greater negativity, as compared with high-frequency names, $F(1, 23) = 4.50$, $p < .05$. In contrast, prominence effects were maximal over left central and posterior electrodes, where prominent names elicited greater negativity than did nonprominent names, $F(1, 23) = 6.09$, $p < .05$. While this difference was not specifically predicted, it likely reflects an asymmetry in the expectation of the critical name across the two conditions. On the nonprominent trials, unlike on the prominent trials, the occurrence of the prime/antecedent can be predicted on half of the trials by the preceding name and conjunction (*Neil and . . .*). There was no additional interaction between frequency and prominence in this region.

Critical name The omnibus ANOVA of the N400 region following the critical name revealed significant main effects of repetition, $F(1, 23) = 12.68$, $p < .01$, and electrode region, $F(6, 138) = 4.16$, $p < .01$, as well as a significant interaction between prominence, repetition, and electrode region, $F(6, 138) = 2.97$, $p < .01$. Subsequent analyses explored these effects in greater detail by separating them with respect to lexical and discourse processing. Analyses of lexical-level processing examined effects of frequency and repetition and focused on nonprominent trials where the use of a repeated name was felicitous (Figs. 4 and 6). The absence of an interaction between repetition and electrode region suggests that priming effects were broadly distributed across all electrodes ($p > .40$). Thus, subsequent analyses collapsed across all regions. Analyses of discourse-level processing focused on effects of frequency and prominence during repeated trials, where the felicity of a repeated name varied with respect to the prominence of its antecedent. The absence of an interaction between prominence and electrode region suggests that priming effects were broadly distributed across all electrodes ($p > .15$). Thus, subsequent analyses collapsed across all regions.

At the lexical level, planned comparisons revealed both a significant main effect of repetition, $F(1, 23) = 6.14$, $p < .05$, and a predicted interaction between repetition and frequency, $F(1, 23) = 3.99$, $p < .05$. Figures 4 and 6 (panels a and b) illustrate that while repetition priming was robust on low-frequency trials, $F(1, 23) = 7.16$, $p < .05$, this effect was not significant on high-frequency trials, $F(1, 23) = 0.58$, $p > .40$. This pattern of results provides new evidence that repeated mention facilitates word recognition the most in cases where episodic memory representations are least robust. There was no additional main effect of frequency (Fs < 1.00, ps $> .30$).

However, at the discourse level, both the main effects of frequency, $F(1, 23) = 0.85$, $p > .30$, and prominence, $F(1, 23) = 0.37$, $p > .50$, and an interaction between frequency and prominence, $F(1, 23) = 2.63$, $p > .10$, failed to reach significance. An additional test was conducted to evaluate whether the repeated name penalty was delayed to the P600 time range, but the omnibus ANOVA (with factors prominence, repetition, frequency, and electrode region) of the P600 range to the critical name revealed only a significant effect of electrode region, $F(6, 138) = 14.35$, $p < .001$. Furthermore, closer inspection of frequency and prominence effects across all regions during repeated trials revealed only that low-frequency names elicited marginally greater positivity than did high-frequency names, $F(1, 23) = 3.06$, $p < .10$. All other effects and interactions failed to reach significance (Fs < 1.00, ps $> .60$). Thus, despite the robustness of a repeated name penalty in other measures, there was no evidence of this effect being reflected in N400 or P600 time ranges after the critical word.

Critical name + 1 The absence of the repeated name penalty in either the N400 or the P600 time window following the critical word suggests that discourse effects in this study may be significantly delayed, emerging later in the processing of downstream text, as is seen in behavioral studies using eye tracking during reading (Ledoux et al., 2007) and neurophysiology studies using ERPs (Kuperberg et al., 2010). To evaluate this possibility, an omnibus ANOVA, focusing on frequency and prominence, was performed on the N400 region to the word following the critical name. The analysis showed significant main effects of frequency, $F(1, 23) = 20.56$, $p < .001$, and electrode region, $F(6, 138) = 5.53$, $p < .01$, as well as significant interactions between frequency and prominence, $F(1, 23) = 7.30$, $p < .05$, frequency and electrode region, $F(6, 138) = 5.01$, $p < .01$, and frequency, prominence, and electrode region, $F(6, 138) = 5.82$, $p < .001$. Planned comparisons revealed both a main effect of frequency, $F(1, 23) = 20.94$, $p < .001$, and an interaction between frequency and prominence that was maximal over left-posterior electrodes,

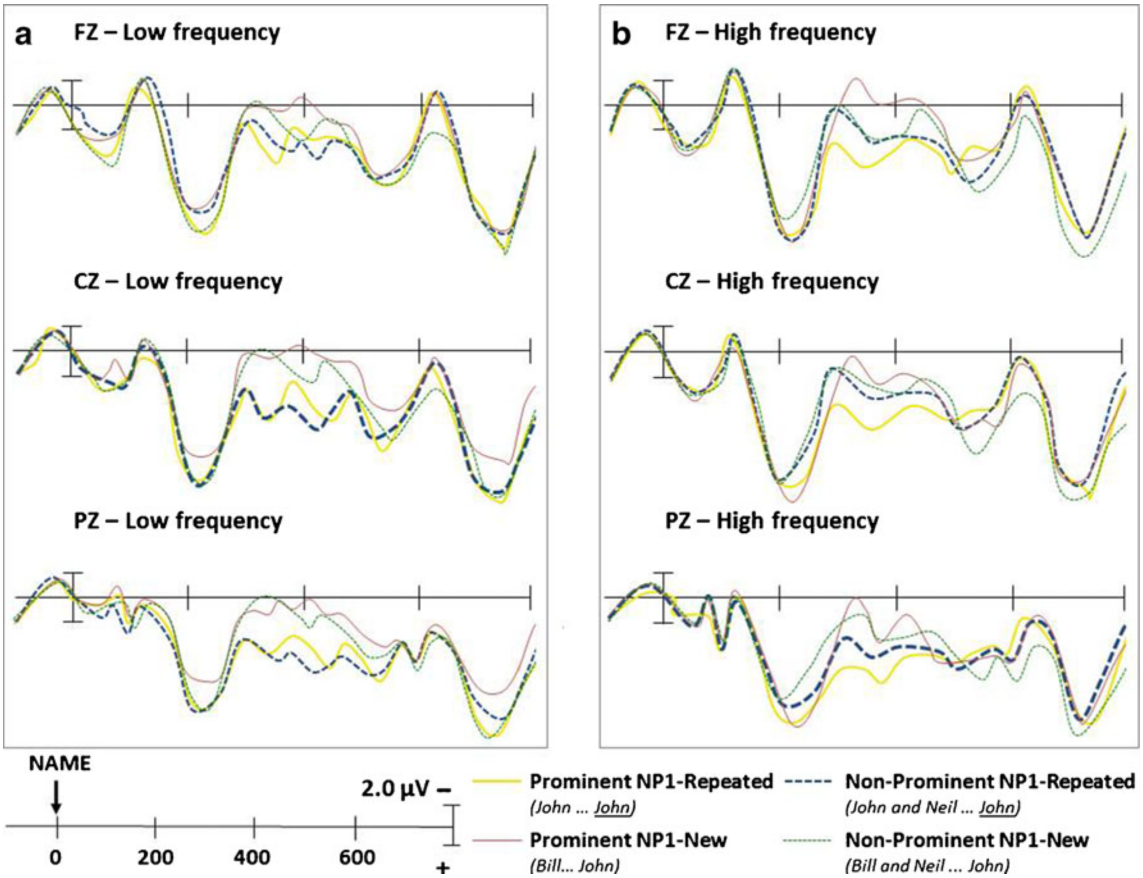

Fig. 4 In Experiment 2, the effect of prominence (prominent vs. nonprominent) on the interpretation of co-referring expressions (repeated vs. new names) in **a** low-frequency and **b** high-frequency conditions. The ERPs were time-locked to the critical name and reflect grand averages across all participants, recorded from midline sites (Fz, Cz, Pz)

$F(1, 23) = 12.88$, $p < .01$. Figures 5 and 6 (panels c and d) illustrate that effects of the repeated name penalty were present among low-frequency names, where repeated names co-referencing prominent antecedents generated more negativity than those co-referencing nonprominent antecedents, $F(1, 23) = 7.83$, $p < .05$. No such difference was found among high-frequency names, $F(1, 23) = 1.93$, $p > .15$. There was no additional main effect of prominence (Fs < 3.00, ps $> .15$).

Since the time-locked word was constant across all conditions, either an auxiliary verb (*was, could*) or the main verb (*talked, wrote*), differences in ERPs cannot be attributed to differences in the eliciting word. Therefore, these effects are likely due to the processing of the critical word, but delayed well beyond the initial N400 or P600 time window. However, the exact interpretation of what these late effects represent needs to take into account the difficulty of determining a true baseline level of activity for these conditions. This will be discussed further below.

Discussion

In Experiment 2, we distinguished between lexical and discourse processes by measuring the modulation of the N400 response following variations in the frequency of names, the prominence of antecedents, and the repetition of co-referring expressions. Consistent with prior results, we found that low-frequency *antecedents* were more difficult to process than their high-frequency counterparts, generating greater N400s (Allen et al., 2003; Van Petten & Kutas, 1990). Critically, we found that the processing of co-referential expressions also varied with frequency. At the lexical level, repetition priming caused a reduced N400 response (to the critical name) that was more exaggerated for low-frequency names, as compared with high-frequency ones. This effect is consistent with previous research demonstrating that lexical access is sensitive to both the inherent statistics of a word and its relational status to other expressions in the sentence (Scarborough et al., 1977; Young & Rugg, 1992). These results suggest that the same signatures of word recognition found in behavioral responses are also present in neural measures. In contrast, at the discourse level, there was no repeated name penalty to the critical name. Instead, this effect was found only much later, in the latency range of the N400 for the following word, where overinformative co-referencing of low-frequency, prominent antecedents generated larger negativities than did co-referencing of nonprominent ones. While this late

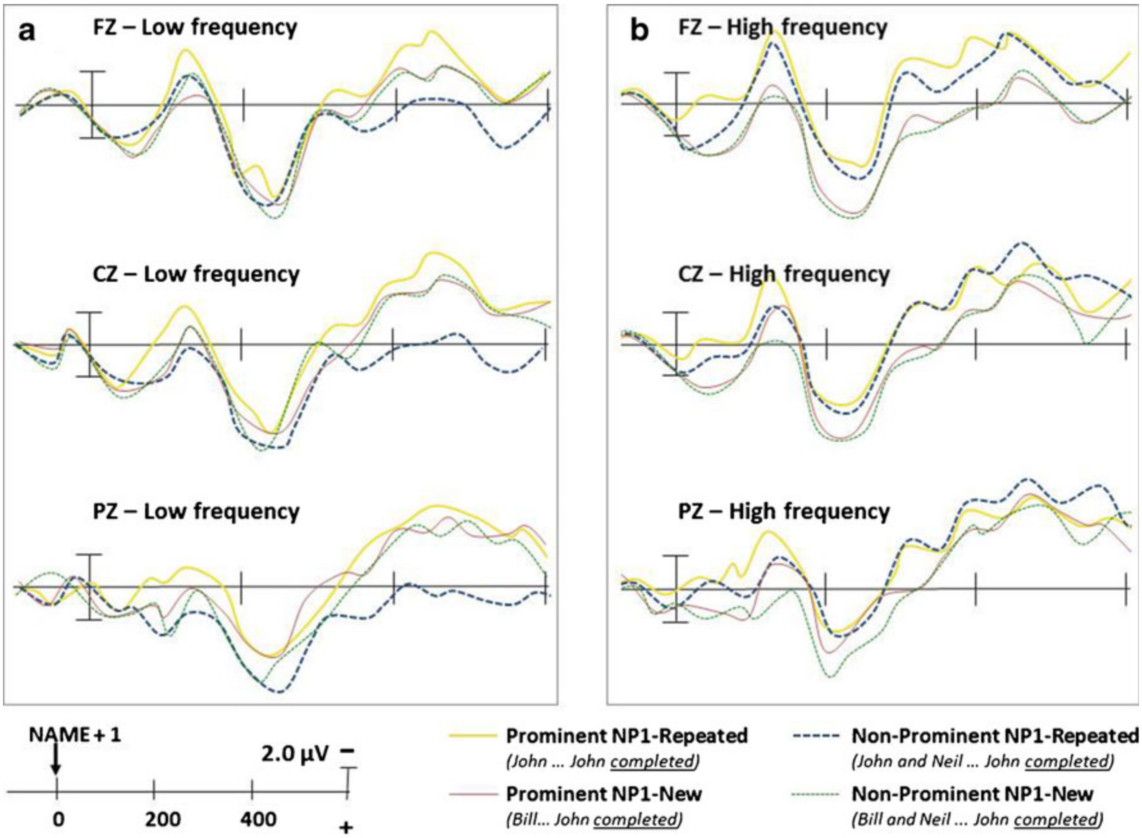

Fig. 5 In Experiment 2, the effect of prominence (prominent vs. nonprominent) on the interpretation of co-referring expressions (repeated vs. new names) in **a** low-frequency and **b**) high-frequency conditions. The ERPs were time-locked to the word *after* the critical name and reflect grand averages across all participants, recorded from midline sites (Fz, Cz, Pz), recorded from frontal (F3, Fz, F4), central (C3, Cz, C4), and posterior (P3, Pz, P4) sites

N400-like effect is notably different from previous ERP demonstrations of the repeated name penalty (Camblin, Ledoux, Boudewijn, Gordon, & Swaab, 2007; Ledoux et al., 2007; Swaab et al., 2004), direct comparisons are difficult given its delay and restriction to low-frequency names.[1] Nevertheless, the presence of divergent time courses between repetition priming and the repeated name penalty suggests that N400s may be produced by multiple neural processes that reflect both lexical access and integration.

It is surprising that a repeated name penalty was not observed for the high-frequency names (i.e., the N400 responses elicited by the word after repeated, high-frequency names that co-referenced prominent antecedents and those that co-referenced nonprominent antecedents were similar). However, collapsing across levels of prominence, high-frequency names as a group generated larger N400 responses than did low-frequency names. This pattern is actually opposite from the frequency difference observed on the antecedent, suggesting that this late-emerging effect reflects lexical integration, rather than access. Furthermore, the negativity generated by the repetition of high-frequency names was similar in space, time, and magnitude to the repeated name penalty seen on the low-frequency prominent trials. This suggests the possibility that the interface between lexical integration and discourse representations is modulated by the familiarity of expressions. In particular, the ubiquity of high-frequency names facilitated recognition at the lexical level, leading to reduced N400s on the antecedent. However, the higher baseline activation of these expressions in discourse representations may also increase the likelihood that subsequent repetition would be considered overinformative, regardless of the prominence of their antecedents.

[1] Similar to previous work (Hagoort et al., 1993; Kuperberg et al., 2010; Osterhout & Holcomb, 1992), the present study examines the effects on critical name + 1, using a baseline that is time-locked to this word. This does introduce the possibility that critical name + 1 baselines are sensitive to manipulations of the critical name. However, the absence of N400 and P600 effects following the critical name suggests that differences across conditions were minimal during the prestimulus baseline window for critical name + 1. Furthermore, the alternative use of critical name baselines would be less ideal, since it would require artifact rejection over a much wider temporal window, which results in lost data and a corresponding increase in noise level for both the critical name and critical name+1 ERPs.

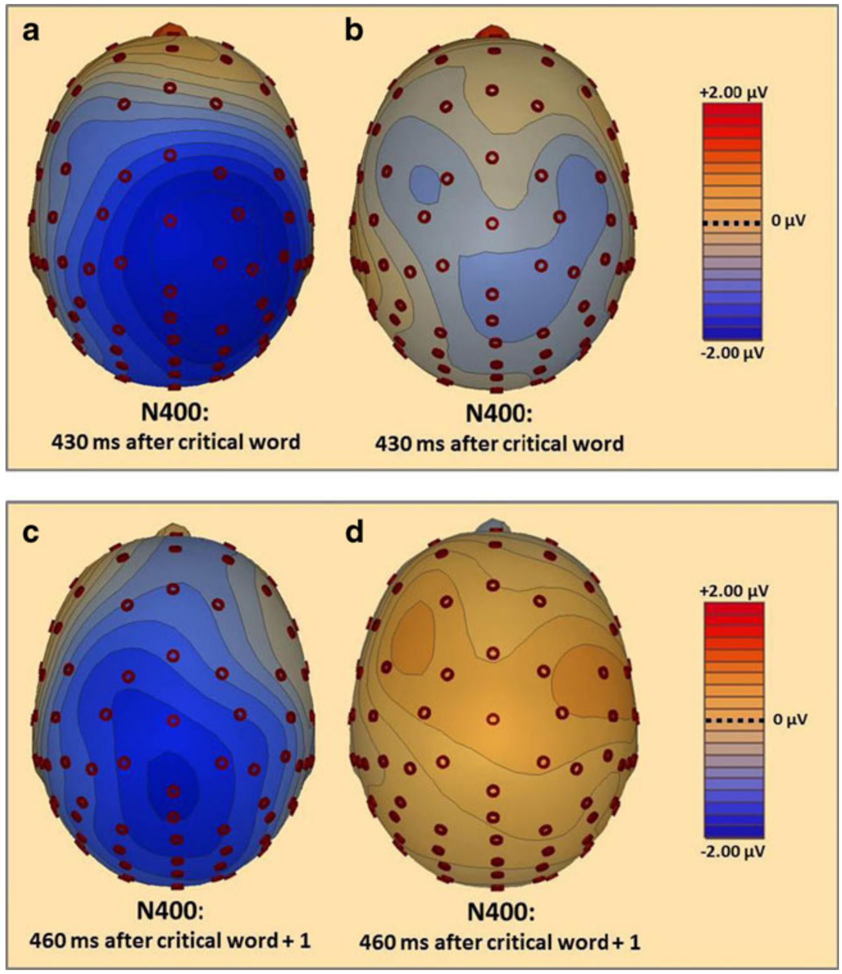

Fig. 6 In Experiment 2, the voltage maps illustrate the spatial distribution of the N400 responses. On **a** low- and **b** high-frequency trials 430 ms after the critical name, N400 responses reflect the new names minus repeated names difference in the nonprominent condition. On **c** low- and **d** high-frequency trials 460 ms after the word following the critical name, N400 responses reflect the prominent antecedents minus nonprominent antecedents difference in the repeated condition

General discussion

Recent research has highlighted the complexity of the mappings between linguistic phenomena, cognitive processes, and neural responses (Kuperberg et al., 2003; Swaab et al., 2004). The present study examined whether overlapping patterns within these mappings could be distinguished through the manipulation of syntactic (Experiment 1) and lexical (Experiment 2) contexts. In two experiments, we found that repeated names facilitated lexical processing, eliciting smaller N400s, as compared with new names. However, relative to this benchmark, evidence of discourse processing was delayed. In Experiment 1, it emerged as a P600 when the repeated name infelicitously co-referenced prominent antecedents from within the same clause. In Experiment 2, it appeared later, after the word following the critical word, as an N400-like effect when repeated names infelicitously co-referenced prominent antecedents from different clauses.

In addition to timing differences, the presence of distinct ERP patterns of interaction with syntactic context provides additional evidence that the lexical and discourse processes are sensitive to separate sources of information. While repetition priming consistently emerged as a N400, the neural responses corresponding to the repeated name penalty varied depending on the structural relationship between the antecedents and referring expressions. This demonstrates that unlike lexical processes, the discourse processes that track co-reference are highly sensitive to the structural properties of sentences. Consistent with prior findings (Gordon et al., 2013), we found that infelicitous repetition of names led to a P600 for co-reference within clauses but a N400 between clauses. Differences in the timing and polarity of these two components suggest the operation of distinct neural processes. One possibility with roots in linguistic analyses of co-reference is that within-clause co-reference is primarily syntactic, while between clauses co-reference is primarily pragmatic. Nevertheless, elaboration and evaluation of this view will depend on a deeper understanding of the types of processes reflected by the N400 and P600, a topic that is the subject of active consideration (e.g., Kuperberg, in press).

These distinct patterns of timing and contextual sensitivity inform our understanding of the interface between the cognitive and neural components of language comprehension. In particular, they suggest that different aspects of meaning are calculated at different levels of interpretation. At the lexical level, visual inputs are incrementally mapped onto orthographic and phonological forms and both are linked to lexical semantics. Consistent with earlier behavioral and neural studies (Scarborough et al., 1977; Young & Rugg, 1992), we found that these processes are sensitive to the inherent statistics of a target word (frequency of *John* vs. *Earl*), as well as its relational occurrence within a sentence (repetition of *John* vs. introduction of *Eric*). At the discourse level, linguistically specified content is matched with a mental model of the event that integrates across past and current context (Gordon & Hendrick, 1998; Huang & Gordon, 2011). In the case of anaphora, the prominence of the antecedent influences the felicity of the expressions used to co-refer. The use of repetition to refer to an already prominent antecedent will be considered overinformative and lead to difficulty in co-referencing.

The present findings also suggest that in addition to being separable procedures, lexical and discourse processes are partially ordered during comprehension. We found that discourse effects followed lexical ones, both when co-referencing occurred within a clause and when it occurred between clauses. These patterns straightforwardly map onto models of comprehension where lexical processes logically precede discourse processes (Ferreira & Patson, 2007; Huang & Gordon, 2011; Tily et al., 2010), although these accounts would still need to contend with prior evidence of interactivity (Kim & Osterhout, 2005; Kuperberg et al., 2003; Kuperberg et al., 2007; Ledoux et al., 2007; Swaab et al., 2004). While a complete synthesis of these patterns is not yet in hand, we believe that the present research suggests a linguistic architecture that involves both functionally separable levels of representations and interactive mappings along their interfaces. It is notable that prior evidence of interactivity involves cases where the relationship between lower- and higher-level representations is highly correlated (e.g., animacy cues from lexical semantics reliably associated with syntactic categories; proper names reliably associated with discourse referents). Critically, unlike models where all available information is used to inform sentence interpretation at its earliest moments (Grodner et al., 2010; Jurafsky, 1996; Levy, 2008), linguistic architectures that build in separable procedures can account for evidence of both cascaded processes and interactivity. In contrast, those that assume a single procedure can account for interactivity but have difficulty explaining evidence of disassociation.

Finally, the present results contribute to recent debates on the nature of the N400. Some accounts have argued that this neural response most directly corresponds to noncombinatorial, lexical access effects (Kutas & Federmeier, 2000; Lau et al., 2008). Evidence in favor of this position comes from neuroimaging data demonstrating similar effects of semantic priming in the posterior temporal cortex (Gold et al., 2006; Helenius, Salmelin, Service, & Connolly, 1998). This region is known for lexical storage, and the combined use of fMRI and MEG methods in these studies reveals neural activation during the 250- to 500-ms window (Lau et al., 2008). Other accounts have suggested that the N400s instead reflect subsequent lexical integration with discourse representations (Brown & Hagoort, 1993; Hauk, Davis, Ford, Pulvermuller, & Marslen-Wilson, 2006). Evidence in favor of this position comes from the time course of this component, relative to even earlier behavior responses. While factors influencing word recognition can reliably affect eye movements 200 ms after stimulus onset (Huang & Gordon, 2011; Ledoux et al., 2007; Liversedge et al., 2003; Raney et al., 2000; Traxler et al., 2000), the peak of the N400 response consistently occurs later than this point (Hauk et al., 2006; Sereno, Rayner, & Posner, 1998).

The findings from Experiment 2 do not distinguish between these two positions. Instead, they provide clear evidence that the N400 responses are elicited by events that are unlikely to reflect a single procedure. Consistent with earlier studies (Ledoux et al., 2007; Swaab et al., 2004), we found that repeated names generated reduced N400 effects (relative to new names) when they co-referenced nonprominent antecedents. This suggests that the N400 reflects facilitation of lexical access. However, repeated names also generated greater N400 effects when they co-referenced prominent antecedents (relative to nonprominent antecedents), although this effect occurred only for low-frequency names and analysis of the delayed N400 required use of baselines that could possibly have been affected by the experimental conditions. This suggests that the N400 reflects effortful integration of overinformative expressions. Critically, while previous studies found no apparent lags in the time course of lexical access and lexical integration, the present work suggests that the two can be distinguished through lexical frequency. We found that for low-frequency names, evidence of lexical *access* continued to emerge on the critical expression. In contrast, evidence of lexical *integration* was delayed until the following word. This suggests that while the lexical access is consistently time-locked to the analysis of the word, lexical integration can appear either on the word or in downstream analysis of subsequent input.

In summary, we explored whether the neural architecture of language comprehension distinguishes between lexical and discourse processes. These experiments assessed the interpretations of co-referential expressions and used the N400 response following repetition priming as a benchmark of lexical processing, with which discourse processing was compared. Manipulations of the syntactic and lexical contexts revealed divergent components and time courses for lexical and discourse processes. These patterns strongly suggest that neural processing of language is sensitive to distinctions in linguistic processes.

Acknowledgments We are grateful to Emily Parks, Patrick Plummer, and Teon Brooks for their assistance in data collection. This work was supported by a National Research Service Award from the National

Institute of Child Health & Human Development to Y.H. (HD061173) and grants from the National Institute of Health to P.C.G. and J.H. (HD060440).

Appendix 1 Examples of stimuli in critical conditions in Experiments 1

1. This morning John/Bill ('s mother) warmly dressed John for school
2. During practice Dylan/Robert ('s niece) did not injure Dylan on purpose
3. While cooking Angie/Lori ('s brother) accidentally cut Angie on her finger
4. After cleaning Fred/Carson ('s girlfriend) wanted to treat Fred to dinner
5. Yesterday Monica/Becky ('s father) spilled popcorn on Monica at the theatre
6. Every morning Frank/Roy ('s girlfriend) enjoyed driving Frank to work
7. Before class Nancy/Eliza ('s dad) makes coffee for Nancy to enjoy
8. Yesterday David/Patrick ('s mom) cleaned the dishes for David without being asked
9. Today Amanda/Elizabeth ('s grandfather) made sandwiches for Amanda on her birthday
10. Today Adam/Ray ('s aunt) granted a break to Adam to rest

Appendix 2 Examples of stimuli in critical conditions in Experiments 2

Low frequency

1. According to the memo Alton/Willis (and Bruce) had plans to research the subject before Alton wrote up the proposal
2. With reluctance (Grant and) Dillon/Kirk washed the dishes while Dillon talked about the up-coming election
3. Each night (Joyce and) Jada/Flora drove downtown because Jada was performing with the symphony orchestra
4. Last week Doris/Peggy (and Cheryl) joined protests against the tuition hike because Doris could not afford the new rate
5. Based on the schedule Clark/Jonas (and Karl) wrote the lyrics to the song before Clark composed the music

High frequency

1. Out in the field Crystal/Lisa (and Lucy) set up the telescope before Crystal started looking at the moon
2. A few days ago John/David (and Albert) went to the post office once John had finished the letters
3. Unfortunately Melanie/Ashley (and Beth) had already left for the ski trip when Melanie caught the flu
4. Despite the distance (Lily and) Morgan/Julia looked for a house near the college after Morgan was mugged downtown
5. Coincidentally (Gary and) Jesse/Aaron had just returned from the library when Jesse mentioned the new book

References

Allen, M. D., Badecker, W., & Osterhout, L. (2003). Morphological analysis during sentence processing. *Language and Cognitive Processes, 18,* 405–430.

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review, 106,* 748–765.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

Camblin, C. C., Ledoux, K., Boudewijn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVPs. *Brain Research, 1146*, 172–184.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

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

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.

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83.

Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language, 33,* 39–68.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience, 26,* 6523–6532.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311–347.

Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic coreference. *Cognition, 62,* 325–370.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22,* 389–424.

Gordon, P.C., Kacinik, N., & Swaab, T.Y. (2013). *Mechanisms of coreference*. Manuscript under review.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some", and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116,* 42–55.

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.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–441.

Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage, 30,* 1383–1400.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain, 121,* 1133–1142.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–62.

Huang, Y., & Gordon, P. (2011). Distinguishing the time-course of lexical and discourse processes through context, co-reference, and quantified expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 966–978.

Huang, Y., & Snedeker, J. (2009a). On-line interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology, 58,* 376–415.

Huang, Y., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in five-year-olds: Evidence from real time spoken language comprehension. *Developmental Psychology, 45,* 1723–1739.

Huang, Y., & Snedeker, J. (2011). 'Logic & Conversation' revisited: Evidence for a division between semantic and pragmatic content in real time language comprehension. *Language and Cognitive Processes, 26,* 1161–1172.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Johnson, M. L., Lowder, M. W., & Gordon, P. C. (2011). The sentence composition effect: Processing of complex sentences depends on the configuration of common versus unusual noun phrases. *Journal of Experimental Psychology. General, 140,* 707–724.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194.

Kennison, S. M., & Gordon, P. C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes, 24,* 229–252.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52,* 205–225.

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kuperberg, G.R. (in press). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In: *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension*. Miller, B., Cutting, L., & McCardle, P. (Eds): Baltimore: Paul Brookes Publishing.

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience, 22,* 2685–2701.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15,* 272–293.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100,* 223–238.

Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Eletrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17,* 117–129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflectsemantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Ladusaw, W. A., & Dowty, D. (1988). Towards a non-grammatical account of thematic roles. In W. Wilkins (Ed.), *Syntax and semantics* (Thematic relations, Vol. 21). London: Academic Press.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory and Cognition, 35,* 801–815.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177.

Liversedge, S. P., Pickering, M. J., Clayes, E. L., & Branigan, H. P. (2003). Thematic processing of adjuncts: Evidence from an eye-tracking experiment. *Psychonomic Bulletin and Review, 10,* 667–675.

Lowder, M. W., Choi, W., & Gordon, P. C. (2013). Word recognition during reading: The interaction between lexical repetition and frequency. *Memory & Cognition, 41*, 738–751.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118,* 155–167.

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

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59,* 494–522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials, elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785–806.

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., & Nichols, 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.

Raney, G. E., Therriault, D. J., & Minkoff, S. R. B. (2000). Repetition effects from paraphrased text: Evidence for an integrated representation model of text representation. *Discourse Processes, 29,* 61–81.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22,* 642–647.

Rugg, M. D., Furda, J., & Lorist, M. (1988). The effects of task on the modulation of event related potentials by word repetition. *Psychophysiology, 25,* 55–63.

Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event related potentials. *Memory and Cognition, 15,* 473–481.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology. Human Perception and Performance, 3,* 1–17.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport, 9,* 2195–2200.

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology. General, 140,* 407–433.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience, 16,* 715–726.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Tily, H., Federenko, E., & Gibson, E. (2010). The time-course of lexical and structural processes in sentence comprehension. *Quarterly Journal of Experimental Psychology, 63,* 910–927.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research, 29,* 581–594.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Learning and Memory, 31,* 443–467.

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,* 657–671.

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

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

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition, 19,* 95–112.

Yang, C. L., Gordon, P. C., Hendrick, R., & Hue, C. W. (2003). Constraining the comprehension of pronominal expressions in Chinese. *Cognition, 86,* 283–315.

Young, M. P., & Rugg, M. D. (1992). Word frequency and multiple repetition as determinants of the modulation of event-related potentials in a semantic classification task. *Psychophysiology, 29,* 664–676.