BRAIN RESEARCH 1597 (2015) 139–158

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

Brain Research

Research Report

The interplay between referential processing and local syntactic/semantic processing: ERPs to written Chinese discourse

Jing Yu[a,1], Yaxu Zhang[a,b,*], Julie E. Boland[c], Lin Cai[a]

[a]Department of Psychology, Peking University, Beijing 100871, China
[b]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China
[c]Department of Psychology, University of Michigan, 525 East University, Ann Arbor, MI 48109-1109, United States

ARTICLE INFO

Article history:
Accepted 4 December 2014
Available online 12 December 2014

Keywords:
Syntactic processing
Semantic processing
Phrase structure
Referential ambiguity
Discourse context
Language comprehension
Event-related brain potentials

ABSTRACT

Two event-related brain potential experiments were conducted to investigate the functional interplay between discourse-level referential processing and local syntactic/semantic processing of phrases. We manipulated both the syntactic/semantic coherence of a noun phrase (NP) and the referential ambiguity of the same NP. Incoherence of the NP elicited a P600 effect in both experiments. Referential ambiguities elicited a sustained negativity (Nref) in a subset of the participants in both experiments. Crucially, among participants showing robust Nref effects to referential ambiguity in the coherent condition, Nref effects were absent when the NP was incoherent. These results provide evidence against theories in which referential processing is functionally independent of local syntactic/semantic processing of phrases. Instead, a local phrase anomaly can block aspects of referential processing concerning ambiguity.

1. Introduction

Language comprehension involves the construction of multiple levels of linguistic representation, including not only syntactic and semantic representations, but also discourse level representations, such as a referential representation that indicates who or what is being talked about (Garnham, 2001; Garrod and Sanford, 1994). In order to understand a sentence such as "My brother came by yesterday", we must engage processing at each of these levels. For example, we recognize the word "*brother*", that "*my brother*" is the noun phrase (NP) syntactic subject, that a brother is the entity who carried out the action, and that "*my brother*" refers to a specific person. How exactly these processes are interrelated is one of the central questions about language comprehension. For example, is the referential processing of "*my brother*" dependent upon having first determined that it is a sentence-initial NP? Theories of sentence processing differ in how and when non-syntactic processes, such as engagement of lexical semantics and discourse context, are used, relative to syntactic processes.

Co-reference is one of the essential elements that not only make a discourse semantically coherent (reflecting specific

*Corresponding author at: Department of Psychology, Peking University, Beijing 100871, China. Fax: +86 10 6276 1081.
E-mail address: yxzhang@pku.edu.cn (Y. Zhang).
[1]Present address: School of Psychology, Southwest University, Chongqing 400715, China.

http://dx.doi.org/10.1016/j.brainres.2014.12.013
0006-8993/

Table 1 – Design and stimulus examples for all four critical conditions in Experiments 1 and 2. Examples are given in Chinese, with English glosses and translations. The critical words are in bold.

(1) *Coherent, referentially unambiguous (CONTROL)*
小明/有/一个/弟弟/和/一个/妹妹。/弟弟/很/胖，/妹妹/很/瘦。/那个/弟弟/昨天/刚来过。
Xiaoming/has/one/brother/and/one/sister./The brother/very/fat,/the sister/very/thin./*nage* (DEM PRON, "that")/**brother**/yesterday/just came by.
(Xiaoming has one brother and one sister. The brother is very fat, the sister is very thin. That **brother** just came by yesterday.)
(2) *Coherent, referentially ambiguous (AMBIGUOUS)*
小明/有/两个/弟弟。/其中/一个/弟弟/很/胖，/另一个/很/瘦。/那个/弟弟/昨天/刚来过。
Xiaoming/has/two/brothers./of/one/brothers/very/fat,/the other/very/thin./*nage* (DEM PRON, "that")/**brother**/yesterday/just came by.
(Xiaoming has two brothers. One brother is very fat, the other is very thin. That **brother** just came by yesterday.)
(3) *Incoherent, referentially unambiguous (INCOHERENT)*
小明/有/一个/弟弟/和/一个/妹妹。/弟弟/很/胖，/妹妹/很/瘦。/那个/很/弟弟/昨天/刚来过。
Xiaoming/has/one/brother/and/one/sister./The brother/very/fat,/the sister/very/thin./*nage* (DEM PRON, "that")/*hen* ("very")/**brother**/yesterday/ just came by.
(Xiaoming has one brother and one sister. The brother is very fat, the sister is very thin. That very **brother** just came by yesterday.)
(4) *Incoherent, referentially ambiguous (DOUBLE)*
小明/有/两个/弟弟。/其中/一个/弟弟/很/胖，/另一个/很/瘦。/那个/很/弟弟/昨天/刚来过。
Xiaoming/has/two/brothers./of/one/brothers/very/fat,/the other/very/thin./*nage* (DEM PRON, "that")/*hen* ("very")/**brother**/yesterday/just came by.
(Xiaoming has two brothers. One brother is very fat, the other is very thin. That very **brother** just came by yesterday.)

DEM PRON, demonstrative pronoun.

nature of connections between concepts or facts) but also underlay the establishment of conceptual cohesion (reflecting global fit of concepts) (see Barton and Sanford, 1993; Sanford and Garrod, 1989). Later sentences refer back to people, objects, and events introduced in earlier sentences. For example, consider the discourse in (1) in Table 1. A brother is introduced in the first sentence and referred back to again with the noun phrases (NPs) "*the brother*" and "*that brother*", respectively. We will call these co-referential expressions in the second and third sentences 'repeated noun anaphors', as in Cloitre and Bever (1988). In (1), there is a unique referent for these anaphors, but in (2) in Table 1, the same repeated noun anaphor in the third sentence is ambiguous, because two brothers were introduced in the first part of the discourse. In such a case, the ambiguity is often resolved with a prenominal or postnominal modifier, as in "that very fat brother just came by yesterday". In this paper, we examine referential processing in ungrammatical sentences, illustrated in (3) and (4) in Table 1, in which the anaphoric noun phrase is both syntactically and semantically incoherent because there is an intensifier without an adjective. Note that the NP is syntactically and semantically anomalous regardless of whether it is in a two-brother context (4) or a one-brother context (3).

In these examples, the referring expression is the full NP, i.e., "*that brother*" in (1) and "that very fat brother". If an incoherent NP blocks referential processing, the referential ambiguity in (4), compared with its unambiguous counterpart in (3), would not be processed. We call this the Blocking Hypothesis. A second possibility is that the syntactic and semantic problem in the repeated noun anaphor would be detected and repaired prior to referential interpretation. The repair would require either deleting "*very*", along with any adjectival phrase constituent that was built to attach "*very*" to the demonstrative pronoun "*that*", or adding an adjective between "*very*" and "*brother*". We call this the Repair Hypothesis, and it predicts that processing of the referential ambiguity in (4) would be delayed, relative to (2). A third possibility is that referential processing does not depend upon having encountered a well-formed NP; referential processing can be triggered by incomplete or malformed NPs. We call this the Independence Hypothesis, because it assumes that referential processing does not depend upon having complete, well-formed syntactic and semantic representations.

In the next section, we will describe some foundational work establishing the Nref as a reliable index of the processing of referential ambiguity within event-related potential (ERP) experiments. Then we will consider our three hypotheses in turn, together with the empirical evidence.

1.1. *The Nref effect*

In the current study, we focus on the referential aspect of discourse processing—i.e., identifying a suitable discourse referent for a NP like "*that brother*". Following the pioneering work of van Berkum and colleagues, we will be using the Nref effect to measure the processing of referential ambiguity (e.g., Nieuwland et al., 2007; Nieuwland and van Berkum, 2006, 2008a; van Berkum et al., 1999, 2003; see Nieuwland and van Berkum, 2008b; van Berkum, 2004; van Berkum et al., 2007, for review). The Nref is a frontal, sustained negative shift that emerges about 300 ms after the onset of a referentially ambiguous noun, such as the bolded "*girl*" in (5a), relative to its unambiguous counterpart, such as the bolded "*girl*" in (5b) (van Berkum et al., 1999).

> (5a) David had told the two girls to clean up their room before lunch time. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time. David told the girl ….
> (5b) David had told the boy and the girl to clean up their room before lunch time. But the boy had stayed in bed all morning, and the girl had been on the phone all the time. David told the girl ….

In addition to referential ambiguity, an Nref-like effect has also been observed for a mismatch in syntactic gender when the problematic expression had referential implications. Martin et al. (2012, 2014) found such an effect in sentences like the Spanish version of (6), where the determiner "*another*" in a noun-phrase ellipsis construction does not match the antecedent "*t-shirt*" in gender. The gender mismatch between "*another*" and its antecedent creates a referentially problematic expression rather than an ambiguity, but the two situations are actually similar. In both (5a) and (6), the Nref effect occurred at a point in the sentence at which there should have been (but actually was not) a clear one-to-one mapping between a referring expression and its antecedent.

(6) Marta bought the t-shirt (fem.) that went well with the dress (masc.) and Miren took another (masc.) ….

Unfortunately, a robust Nref effect is not necessarily observed for all participants: in Nieuwland and van Berkum (2008a), 19 of 39 participants exhibited a late positivity rather than an Nref effect for referentially ambiguous NPs. This late positivity is presumed to be a type of P600 effect. Although a P600 is often found for syntactic anomalies, other types of anomalies can elicit P600s as well. In fact, a referentially induced P600 effect has been found (in the full set of participants) for pronouns without a referent (e.g., Nieuwland and van Berkum, 2006) and for reflexive pronouns that disagree in gender with their antecedent (Osterhout et al., 1997). Together, these findings indicate that, while referential ambiguity often elicits an Nref effect, referentially failing or gender-mismatching pronouns elicit a P600 effect, and in some participants, the referential ambiguity effect and referential failing or gender-mismatching effect may be indistinguishable.

Although the Nref effect first appears at about the same time as another negative component, the semantic N400, the Nref differs in that it is sustained over half a second or more without a well-defined peak, and tends to be more frontal in scalp distribution (largest and particularly persistent over anterior sites, and somewhat larger over the left than over the right hemisphere). In the first report of the Nref effect, the mean magnitude of the effect was 0.8 μV and the effect did not significantly interact with electrode site in the 300–600-ms time window (van Berkum et al., 1999).

Interestingly, referentially ambiguous nouns did not evoke an Nref effect when the referentially ambiguous noun was semantically anomalous, as is "*necklace*" in the Dutch version of (7) (Nieuwland and van Berkum, 2008a). Such nouns elicited an N400 effect, but no Nref effects, suggesting that a local semantic anomaly can preclude more global discourse referential processing concerning ambiguity (anaphoric inferencing). In addition, there were no Nref effects for the nouns that were ambiguous in terms of recently activated memory tokens in the textbase but were actually referentially unambiguous within the situational context, as is the bolded "*nephew*" in the last sentence of the Dutch version of (8) (Nieuwland et al., 2007). This finding suggests that the Nref effect reflects genuine, situation-model ambiguity, rather than superficial ambiguity. Thus, for those participants who produce a clear Nref effect, it can be used to track certain aspects of referential processing at a deep, situation-model level (for further discussion about the functional interpretation of the Nref, see Nieuwland and van Berkum, 2008b; van Berkum et al., 2007).

(7) Britney Spears had several pieces of jewelry, including a golden necklace and a silver one. One day she was about to leave for a gala. She stepped into the necklace ….

(8) At the family get-together, Jim had been talking to one nephew who was very much into politics and another one who was really into history. But Jim himself was only interested in sports, cars, girls etc. The nephew who was into history left early, but the nephew who was into politics kept rambling on. Jim didn't understand one bit and got rather bored. He told the nephew ….

Like Nieuwland and van Berkum (2008a), we will use the Nref as an index of referential processing, and more specifically, as evidence that a referential ambiguity has been processed. And as in their study, we will focus our analysis on those participants for whom the referential ambiguity evokes a clear Nref effect. Our primary goal is to test the three hypotheses using discourse materials exemplified in Table 1. Like English, Mandarin has a fixed word order within the NP, and a noun following a word like "*very*" ("*hen*") is both syntactically and semantically anomalous.

1.2. *The blocking hypothesis*

The Blocking Hypothesis is motivated by theoretical accounts that presume referential processing to depend upon successful syntactic processing, successful semantic processing, or both. Thus, the Blocking Hypothesis is consistent with models that assume a functional primacy of syntactic processing over referential processing, such as versions of the extended argument-dependency model or eADM (Bornkessel and Schlesewsky, 2006; Bornkessel-schlesewsky and Schlesewsky, 2008, 2009)[2]. In these models, phrase structuring using syntactic category information is a prerequisite for semantic and discourse context processing either in a strictly serial or in a cascaded fashion. The Blocking Hypothesis is also consistent with theories in which anaphoric inferences associated with referential ambiguity are disrupted by semantic incoherence (Nieuwland and van Berkum, 2008a). Nieuwland and van Berkum developed such an account, based on the assumption that limited processing resources would be directed to the most salient disruptions, with semantic (and presumably also syntactic) disruptions constituting a greater threat to coherence than referential ambiguity.

1.3. *The repair hypothesis*

The Repair Hypothesis is essentially a softer version of the Blocking Hypothesis. The assumption is that some types of syntactic and semantic anomalies can be diagnosed and repaired (e.g., Fodor and Inoue, 1994). In such cases, the same syntax-first and semantics-first theories that were used to motivate the Blocking Hypothesis would predict that referential processing is not blocked completely, but merely delayed by some types of syntactic and semantic incoherence. Correspondingly, processing of the referential ambiguity would be delayed in (4) compared with (2) from Table 1, and the Nref effect would also be delayed.

[2]Current formulations of this theory maintain that processing proceeds in a cascaded fashion, so this account might also be consistent with the Independence Hypothesis, depending upon how the parameters are set for cascading between different types of processing.

1.4. The independence hypothesis

One possibility is that referential processing is independent of lower level processing, such that referential processes proceed whether or not an NP is complete or well-formed (see Nieuwland and van Berkum, 2008a for a discussion about the possible independence of referential processing from semantic processing). There is already evidence for referential interpretations of incomplete NPs, as predicted by the Independence Hypothesis. One of the most striking features of sentence comprehension is its incremental nature. Readers and listeners show evidence of interpretation well before all temporary ambiguities are resolved, and this is true in terms of referential processing as well as syntactic and semantic processing (e.g., Altmann and Steedman, 1988; Boland and Blodgett, 2001; Tanenhaus et al., 1995; Trueswell and Tanenhaus, 1991; van Berkum et al., 1999). Such anticipatory processing has been especially apparent in visual world eye tracking experiments that measure looks to potential referents of spoken NPs (e.g., Hanna and Brennan, 2007; Sedivy et al., 1999; Weber et al., 2006). For example, Sedivy and colleagues found that listeners who heard "Pick up the tall …" could use a visual contrast set (two objects of the same type, that differed in tallness) to interpret the scalar adjective prior to hearing the head noun. That is, even though there were two tall objects in the display, listeners looked at the one that was a member of a contrast set. Weber et al. (2006) found that this effect could be modulated by contrastive prosody. Findings like these demonstrate that a full NP is not always necessary for the initiation of referential processing, which can exploit a rich variety of cues beyond the explicit linguistic content.

Despite the evidence for referential processing of incomplete NPs, referential processing could still be blocked by a syntactic or semantic anomaly. In the current study, we provide a stronger test of the Independence Hypothesis by investigating whether referential processing could proceed over a syntactically and semantically incoherent NP, such as "that very brother". Would the referential ambiguity be processed, without the need for syntactic and/or semantic repair processes? In order to maximally contrast the Independence Hypothesis with the Blocking/Repair Hypothesis, we take the Independence Hypothesis to predict that referential processing proceeds for NPs containing syntactic and semantic incoherence. This hypothesis is consistent with Hagoort's (2003, 2005) unification model, in which referential computations can occur before local syntactic/semantic processing of phrases is completed.

In considering the Independence Hypothesis, we assume that the demonstrative pronoun "*that*" projects either a NP or a determiner phrase (DP). Regardless of the theory-internal details of this structure, a NP must be projected to attach "*very*" to "*that*", and the NP node would provide a potential attachment site for the noun "*brother*". Because the modifier "*very*" doesn't provide any disambiguating information, referential processing sufficient to determine whether the malformed NP has a unique referent is not possible until the noun "*brother*" is encountered. However, at the critical noun, the Independence Hypothesis predicts that an Nref should occur if "*that brother*" does not have a unique referent.

1.5. The current study

We designed this study to determine whether referential processing is blocked, delayed, or otherwise impacted when the referring expression is syntactically and semantically incoherent. We therefore used a full factorial design that crossed the coherence of the NP (coherent vs. incoherent) with referential ambiguity (ambiguous vs. unambiguous) of the same NP. Stimuli were three-sentence discourses, as illustrated in Table 1.

The subject noun of the third sentence in each short discourse (*brother*) served as the critical word. The critical word was coherent and referentially unambiguous in the CONTROL condition. For the referentially ambiguous only (AMBIGUOUS) condition, the critical word had two equally eligible candidate referents. For the incoherent only (INCOHERENT) condition, the degree adverb *hen* ('*very*') was inserted immediately before the critical noun, resulting in syntactic and semantic incoherence of the NP. In Chinese, *hen* can be followed by an adjective or a verb but not by a noun (Lü and Zhu, 1979). The semantic incongruence was due to the degree adverb *hen* being followed by a noun that cannot be gradable. The same manipulation was also used in Zhang et al. (2010). In the incoherent and referentially ambiguous (DOUBLE) condition, the critical noun had two equally eligible referents and *hen* appeared immediately before it.

We expected the AMBIGUOUS condition to elicit an Nref effect beginning about 300 ms after the critical word (e.g., Nieuwland et al., 2007; Nieuwland and van Berkum, 2006, 2008a; van Berkum et al., 1999, 2003), as described above. For the INCOHERENT condition, we expected a P600 effect, which is frequently found for syntactic incoherence (for review, see Osterhout et al., 2004), including syntactic category incoherence of the type used in the present study (*hen*+noun) (Zhang et al., 2010).

More importantly, consider the predicted data pattern if the referential processing of the critical noun necessarily depends upon successful syntactic processing, successful semantic processing, or both, as assumed in the Blocking Hypothesis and the Repair Hypothesis. The inability to construct a coherent local phrase for the incoherent NPs would block or delay referential interpretations of these NPs, similar to the scenario observed by Nieuwland and van Berkum (2008a) in their investigation of the interplay between local semantic integration and referential ambiguity processing. In such a case in which blocking or repair occurs, the ERPs elicited by the critical nouns for the DOUBLE condition should be the same as those for the INCOHERENT condition during the earliest time window in which an Nref effect is observed, because the referential ambiguity (Nref) effect in the DOUBLE condition would be absent, due to the incoherence of the NP. The Blocking and Repair Hypotheses both predict an interaction of referential ambiguity with coherence across the four conditions, because referential effects would be observed only when the NP is coherent.

In contrast, if a well-formed NP is not always necessary for referential processing, as assumed in the Independence Hypothesis, then referential processing may proceed in the absence of successful construction of the NP. If so, an immediate referential ambiguity (Nref) effect would occur even when the NP is syntactically and/or semantically incoherent. Only the Independence Hypothesis can account for a difference between the DOUBLE and INCOHERENT conditions as soon as the Nref effect emerges in the AMBIGUOUS condition, in contrast to the Blocking and the Repair Hypotheses.

Clearly, we can only test these hypotheses if participants exhibit an Nref effect to referential ambiguity. Thus, we will follow Nieuwland and van Berkum's (2008a) strategy of

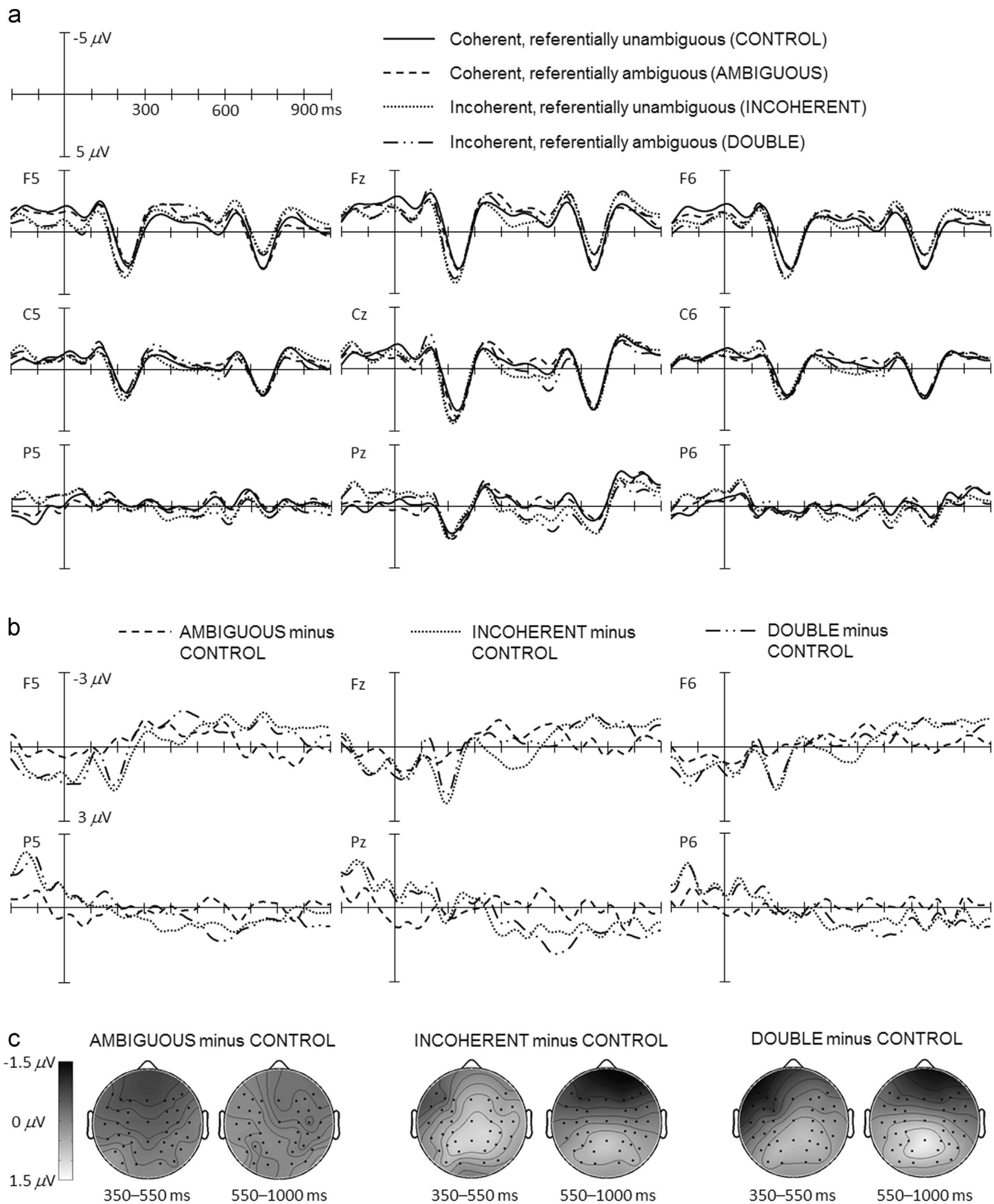

Fig. 1 – ERPs time locked to the onset of the critical words for all 16 participants in Experiment 1. (a) Grand average ERPs for all four critical conditions at nine scalp sites; (b) three difference waves at six scalp sites; (c) the scalp topographies of the three difference waves in two time windows. In this and the following figures, the onset of the critical words is at 0 ms, and negativity is plotted upwards and waveforms are filtered (10 Hz low pass, 24 dB/oct).

separately analyzing the subset of participants who exhibit the basic Nref effect, if necessary.

We conducted two experiments, which differed mainly in the filler items (see Sections 4.1.2 and 4.2.2 for details) and the task used. In Experiment 1, we used a probe verification task, in which a probe word occasionally appeared after the presentation of the short discourse and participants were asked to judge whether the probe had been present in the preceding discourse. In Experiment 2, a YES/NO comprehension question task was presented after the third sentence in the discourse.

2. Results

2.1. Experiment 1

2.1.1. Behavioral results

The overall average accuracy on the probe verification task was 98.00% (SD=2.37%), equal for all four conditions, suggesting that participants read the stimuli attentively.

2.1.2. ERP results

This section is organized as follows. First we present the grand average ERPs for all four conditions and describe the procedure used to determine the time windows for statistical analysis. Then we report the statistical results for each time window, reporting results from midline sites before results from lateral sites. For all analyses, only effects involving the factors coherence and referential ambiguity are reported.

Fig. 1a shows grand average ERPs elicited by the critical nouns for all four critical conditions, Fig. 1b shows the difference waves of each non-control condition (AMBIGUOUS, INCOHERENT, and DOUBLE) minus CONTROL, and Fig. 1c shows the corresponding scalp topographies of these difference waves in the two main time windows selected for statistical analyses.

As shown in Fig. 1, the INCOHERENT condition elicited an early anterior positivity of short duration, a late anterior negativity, and a typical P600 response, as did the DOUBLE condition. The AMBIGUOUS condition elicited a negativity (Nref) that was not very sustained (we will discuss this after having reported the results of Experiment 2). The DOUBLE condition appears to be more negative than the INCOHERENT condition around the 350–550 ms interval.

In order to more objectively determine the time windows for statistical verification of the observations above, we performed analyses in adjacent 50 ms time windows in the interval starting from 200 ms before and ending 1000 ms after critical word onset, with differences being considered reliable when they were significant in at least two adjacent time windows (Gunter et al., 2000). These and the following ERP analyses were all performed separately for midline and lateral electrodes. Omnibus ANOVAs for midline electrodes included three within-subject factors: electrode (Fz/Cz/Pz), coherence (coherent/incoherent), and referential ambiguity (ambiguous/unambiguous). Omnibus ANOVAs for lateral electrodes included four within-subject factors: hemisphere (left/right), region (anterior/central/posterior), coherence, and referential ambiguity. Crossing the variables of region and hemisphere yielded six regions of interest, with six electrodes for each region of interest: left anterior (F3, F5, F7, FC3, FC5, and FT7), left central (C3, C5, T7, CP3, CP5, and TP7), left posterior (P3, P5, P7, PO3, PO7, and O1), right anterior (F4, F6, F8, FC4, FC6, and FT8), right central (C4, C6, T8, CP4, CP6, and TP8), and right posterior (P4, P6, P8, PO4, PO8, and O2).

The 50-ms interval analyses found that the incoherent conditions were more negative than the coherent conditions in the interval starting from 50 ms before and ending 100 ms after critical word onset at lateral posterior regions, resulting in an interaction of coherence with region for each time window ($F(2,30)=11.31–17.31$, $ps<0.005$, $MSE=1.64–1.92$). These differences should at least partly result from stimuli prior to the presentation of critical word (CW). Note that the word preceding CW differed between the coherent and incoherent conditions (e.g., '*that*' and '*very*', respectively, in Table 1), resulting in a waveform difference between the two conditions that could spill over into the ERPs of CW. In the two adjacent 50-ms time windows in the 150–250 ms interval after CW onset, the incoherent conditions were more positive than the coherent conditions at anterior regions (including Fz), resulting in an interaction of coherence with electrode or with region for each time window ($F(2,30)=11.73–28.41$, $ps<0.005$, $MSE=0.49–1.14$). After that, there were no differences across conditions in two adjacent 50-ms time windows (corresponding to the 250–350 ms interval).

In three adjacent 50-ms time windows in the 350–500 ms interval, the incoherent conditions were more positive than the coherent conditions at the midline sites ($F(1,15)=5.15–6.17$, $ps<0.05$, $MSE=3.56–6.03$). The incoherent conditions were also more positive than the coherent conditions over right anterior-central, right central, or bilateral posterior regions in three adjacent 50-ms time windows in the 400–550 ms interval, resulting in a three-way interaction for each time window ($F(2,30)=7.10–7.60$, $ps<0.01$, $MSE=0.17–0.22$). In addition, the ambiguous conditions were more negative than the unambiguous conditions in the two adjacent 50-ms time windows in the 350–450 ms interval at both the midline and lateral sites ($F(1,15)=4.58–11.48$, $ps<0.05$, $MSE=2.13–3.26$), and in three adjacent 50-ms time windows in the 400–550 ms interval over lateral anterior regions, resulting in an interaction of referential ambiguity with region for each time window ($F(2,30)=4.88–5.81$, $ps<0.05$, $MSE=0.64–0.93$).

In the 550–1000 ms interval, the incoherent conditions were more positive than the coherent conditions in each 50-ms time window (except for the 850–900 ms window) over posterior regions (including Pz), reflected in an interaction of coherence with electrode or with region for each of these time windows ($F(2,30)=5.02–30.96$, $ps<0.05$, $MSE=0.70–2.70$). In addition, the incoherent conditions were more negative than the coherent conditions over anterior regions (including Fz) in five adjacent 50-ms time windows in the 650–900 ms interval, reflected in an interaction of coherence with electrode or with region for each time window ($F(2,30)=8.20–30.96$, $ps<0.01$, $MSE=0.75–2.30$).

Three time windows were chosen on the basis of the results of these 50-ms interval analyses and earlier studies (e.g., van Berkum et al., 1999; Zhang et al., 2010): (a) 150–250 ms after CW onset for early, anterior (including Fz) positivity effects to incoherence of the NP; (b) 350–550 ms for the negativity effects to referential ambiguity (this time window would also cover the time range of the broad positivity effects to incoherence), and (c) 550–1000 ms for posterior (including Pz) positivity effects to incoherence (this time window would also cover the time range

Table 2 – Experiment 1 overall analyses of variance for three time windows (in milliseconds) for midline and lateral electrodes ($N=16$).

Source	*dfs*	150–250			350–550			550–1000		
		F	*p*	*MSE*	*F*	*p*	*MSE*	*F*	*p*	*MSE*
Midline electrodes										
COH	1,15	3.63	0.076	6.04	8.12	0.012	3.58	2.07	0.171	1.03
REF	1,15	<1			8.62	0.010	1.53	<1		
COH × electrode	2,30	19.28	<.0005	0.46	1.12	0.314	1.44	38.83	<.0005	0.44
REF × electrode	2,30	<1			3.61	0.067	0.47	2.58	0.113	0.21
COH × REF	1,15	1.55	0.232	3.02	<1			1.50	0.240	1.53
COH × REF × electrode	2,30	<1			<1			<1		
Lateral electrodes										
COH	1,15	1.97	0.181	6.08	6.69	0.021	1.91	<1		
REF	1,15	<1			8.92	0.009	1.27	1.23	0.284	1.19
COH × region	2,30	35.53	<.0005	0.47	2.85	0.109	1.66	40.38	<.0005	0.48
COH × hemisphere	1,15	<1			8.56	0.010	0.51	<1		
COH × region × hemisphere	2,30	<1			9.01	0.002	0.13	<1		
REF × region	2,30	<1			7.55	0.010	0.45	<1		
REF × hemisphere	1,15	1.37	0.260	0.29	1.31	0.270	0.16	2.12	0.166	0.36
REF × region × hemisphere	2,30	<1			1.74	0.206	0.08	3.93	0.042	0.04
COH × REF	1,15	<1			1.18	0.295	0.82	1.01	.330	1.32
COH × REF × region	2,30	<1			<1			<1		
COH × REF × hemisphere	1,15	<1			<1			<1		
COH × REF × region × hemisphere	2,30	<1			<1			<1		

COH=coherence; REF=referential ambiguity.

of anterior negativity effects to incoherence). The results of the overall ANOVAs are shown in Table 2.

2.1.2.1. *The 150–250-ms time window.* As shown in Table 2, at the midline sites, there was an interaction of coherence with electrode. Separate analyses found a larger positivity for the incoherent compared to the coherent nouns at Fz ($F(1,15)=11.44$, $p=0.012$, $MSE=1.22$). For the lateral sites, there was an interaction of coherence with region, which was due to a larger anterior positivity for the incoherent compared to the coherent nouns ($F(1,15)=12.43$, $p=0.009$, $MSE=0.66$). The early positivity, which was also observed for low-expectancy words in sentences (Ferretti et al., 2007), may reflect either more attention or more difficult lexical processing for the incoherent compared to the coherent nouns.

2.1.2.2. *The 350–550-ms time window.* At the midline sites, there was an effect of coherence, with the incoherent nouns being more positive than the coherent nouns. In addition, there was an effect of referential ambiguity, with the ambiguous nouns being more negative than the unambiguous nouns. This is the expected Nref effect.

At the lateral electrodes, there was an effect of coherence, an interaction of coherence with hemisphere, and an interaction among coherence, region, and hemisphere. Separate analyses restricted to each region found an effect of coherence at posterior region, with a larger positivity (P600) for the incoherent compared to the coherent nouns ($F(1,15)=5.06$, $p=0.040$, $MSE=1.12$). At anterior and central regions, there was an interaction between coherence and hemisphere (anterior, $F(1,15)=11.50$, $p=0.004$, $MSE=0.20$; central, $F(1,15)=9.49$, $p=0.008$, $MSE=0.08$), which was due to a larger positivity for the incoherent compared to the coherent nouns at right anterior-central sites (right anterior, $F(1,15)=8.16$, $p=0.024$, $MSE=0.18$; right central, $F(1,15)=18.23$, $p=0.002$, $MSE=0.20$). In addition, there was an effect of referential ambiguity and an interaction of referential ambiguity with region. Separate analyses showed a larger negativity for the ambiguous compared to the unambiguous nouns at anterior region ($F(1,15)=16.11$, $p=0.003$, $MSE=0.17$).

Although referential ambiguity did not interact with coherence, planned comparisons (the least conservative statistical test) were conducted to confirm that there was no difference between the DOUBLE and INCOHERENT conditions at any midline electrode or lateral region ($F<1$ or $p>0.09$). In contrast, a larger negativity (Nref) for the AMBIGUOUS condition compared with the CONTROL condition was found at both Fz ($F(1,15)=12.98$, $p=0.009$, $MSE=0.30$) and lateral anterior region ($F(1,15)=25.26$, $p<0.0005$, $MSE=0.08$).

2.1.2.3. *The 550–1000-ms time window.* At the midline sites, there was an interaction of coherence with electrode. Separate analyses found a larger negativity at Fz and a larger positivity (P600) at Pz for the incoherent compared to the coherent nouns (Fz, $F(1,15)=13.91$, $p=0.006$, $MSE=0.26$; Pz, $F(1,15)=24.84$, $p<0.0005$, $MSE=0.33$).

At the lateral electrodes, there was an interaction of coherence with region. Separate analyses found a larger anterior negativity and posterior positivity for the incoherent compared to the coherent nouns (anterior, $F(1,15)=13.46$, $p=0.006$, $MSE=0.18$; posterior, $F(1,15)=27.97$, $p<0.0005$, $MSE=0.11$). In addition, there was an interaction among referential ambiguity, region, and hemisphere. However, separate analyses restricted to each hemisphere found neither an effect of referential ambiguity nor its interaction with region ($F<1$ or $p>0.09$). Although separate analyses restricted to each region found a marginal interaction of referential ambiguity with hemisphere at central region

($F(1,15)=4.26$, $p=0.057$, $MSE=0.07$), there were no effects of referential ambiguity at either hemisphere (left, $F(1,15)=4.27$, $p=0.112$, $MSE=0.15$; right, $F<1$).

2.1.3. Discussion

Using three-sentence discourses, we constructed four versions of each item by manipulating a noun phrase in the third sentence. One version was completely correct at the critical noun (CONTROL). In one version, the noun was a word of the wrong syntactic and semantic type (INCOHERENT). In one version, the noun was referentially ambiguous (AMBIGUOUS), and in the final version, the noun was both syntactically/semantically incoherent and referentially ambiguous (DOUBLE). Qualitatively different ERP effects were observed for the local incoherence and referential ambiguity, as expected, based on previous studies. The local incoherence elicited a posterior positivity (P600), as in many previous studies (e.g., Friederici et al., 1999; Hagoort et al., 2003; Zhang et al., 2010). In the literature, P600 effects have been correlated with syntactic processing difficulty caused by phrase structure anomaly (for review, see Osterhout et al., 2004), and more recently, with non-syntactic anomalies as well (see Bornkessel-schlesewsky and Schlesewsky, 2008; van de Meerendonk et al., 2009). Besides the P600, incoherent nouns also elicited an early anterior positivity of short duration and a late anterior negativity that was also observed in a previous study (Zhang et al., 2010) and might reflect difficult syntactic/semantic processing for incoherent nouns.

In contrast, referentially ambiguous nouns elicited a negativity (Nref) with a broad distribution at midline sites and an anterior distribution at lateral sites in the 350–550-ms range. The Nref effects were not very sustained, as reflected in an absence of referential ambiguity effects in the 550–1000-ms range. Except for not being very sustained, the Nref effects obtained in the present study are very similar to those observed in previous studies (e.g., van Berkum et al., 1999; see Nieuwland and van Berkum, 2008b; van Berkum, 2004; van Berkum et al., 2007, for review).

The finding of the syntactic/semantic anomaly effects (the P600) together with the referential ambiguity effects (the Nref) is very important, because it provides direct evidence that our participants were building local phrase structure and establishing reference at a deep level, in the context of a probe verification task. This allowed us to evaluate the Blocking, Repair, and Independence hypotheses. In the 350–550-ms window, we found the predicted Nref effect. Although referential ambiguity did not interact with coherence, the effect appeared to be attenuated in the incoherent conditions. Indeed, a planned comparison found no Nref effects for the DOUBLE condition compared to the INCOHERENT condition, so we suspect that the interaction would reach significance if we could reduce the noise level in the experiment.

We therefore conducted Experiment 2 with another set of participants and attempted to address three other limitations of Experiment 1. First, most (210/250) of the short discourses (including the fillers) contained highly similar permutations of a small set of constructions. Second, a probe word verification task was used and the additional memory demands imposed on participants might interact in unknown ways with the process of interest. Third, DC recordings were used and thus no high-pass filter was applied during data acquisition, resulting in a relatively high level of noise due to slow voltage shifts.

To overcome these limitations, in Experiment 2, a larger number of participants were involved, more filler items of various syntactic structures were included, and a comprehension question task was used instead of probe word verification. In addition, EEG data were recorded with a 0.05–100 Hz band-pass filter in order to reduce the level of noise.

2.2. Experiment 2

2.2.1. Behavioral results

The overall average accuracy on the YES/NO comprehension question task was 90.23% (SD=6.60%), equal for all four conditions, suggesting that participants read the stimuli attentively.

2.2.2. ERP results

In this section, we first report the ERP results of all participants of Experiment 2. However, the most important results in this section feature the combined analysis of data from Experiments 1 and 2, divided into those participants who produced a positivity to referential ambiguity and those who did not. For these results, we first describe the motivation for combining experiments and the procedure for dividing participants into the two groups. The results for participants who did not produce a positivity to ambiguity (the no-positivity-to-ambiguity group) are presented first, and the section concludes with the results from the participants who produced a positivity, rather than an Nref, to referential ambiguity. For each group, we first describe the procedure or basis for determining the time windows, and then present the results from each time window. As in Experiment 1, results from midline sites were always reported first, followed by results from lateral sites, for each selected time window.

2.2.2.1. All participants. This section is organized as follows. First we present the grand average ERPs from all four conditions and describe the procedure for selecting the time windows for analysis. Then we report the results of statistical analyses performed on the mean amplitudes in the selected time windows.

Fig. 2a shows grand average ERPs elicited by the critical nouns for all four critical conditions, Fig. 2b shows the difference waves of each non-control condition (AMBIGUOUS, INCOHERENT, and DOUBLE) minus CONTROL, and Fig. 2c shows the corresponding scalp topographies of these difference waves in the two main time windows selected for statistical analyses, for all 24 participants in this experiment.

As in Experiment 1, there was an early, more anterior positivity response of short duration and a typical P600 effect for both the INCOHERENT and DOUBLE conditions. However, in contrast to Experiment 1, the AMBIGUOUS condition did not elicit obvious negativity (Nref) effects. Instead, a late positivity was evoked. As noted above, our hypotheses assume an Nref in the AMBIGUOUS condition. In addition, participants who showed a late posterior positivity to referential ambiguity (Nieuwland and van Berkum, 2008a) or a broad positivity to referentially problematic

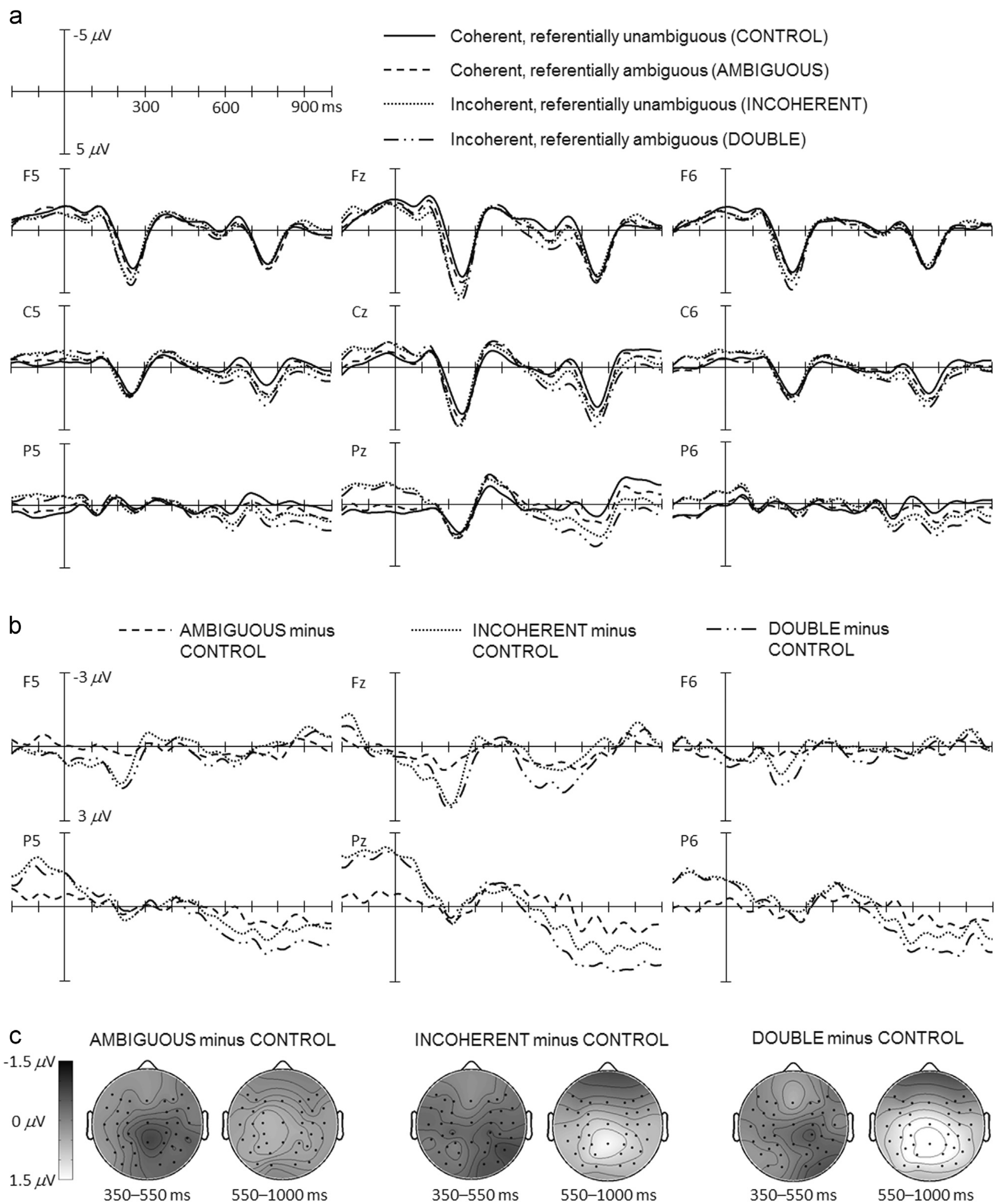

Fig. 2 – ERPs time locked to the onset of the critical words for all 24 participants in Experiment 2. (a) Grand average ERPs for all four critical conditions at nine scalp sites; (b) three difference waves at six scalp sites; (c) the scalp topographies of the three difference waves in two time windows.

expressions due to gender mismatch between determiner and antecedent (Martin et al., 2014) have been reported not to show an Nref to referential ambiguity or retrieval interference during the building of determiner-antecedent dependency. Therefore, we will follow two previous studies (Martin et al., 2014; Nieuwland and van Berkum, 2008a) and conduct a separate analysis using the subset of participants who did not exhibit a late positivity in the AMBIGUOUS condition (i.e., the no-positivity-to-ambiguity group, see Section 2.2.2.2). However, we first report an analysis using the complete dataset. For this analysis, in order to more objectively determine the time windows for statistical verification of the observations above, we performed analyses in adjacent 50 ms time windows in the interval starting from 200 ms before and ending 1000 ms after critical word (CW) onset, with differences being considered reliable when they were significant in at least two adjacent time windows, as in Experiment 1.

The 50-ms interval analyses found that the incoherent conditions were more negative than the coherent conditions at centro-posterior regions (including Cz and Pz) in the interval starting from 50 ms before and ending 100 ms after CW onset, resulting in an interaction of coherence with electrode or with region for each time window ($F(2,46)=21.02$–45.93, $ps<0.0005$, $MSE=0.74$–1.48). These differences must be, at least partly, a response to stimuli prior to the presentation of CW. As mentioned in Section 2.1.2, the word preceding CW differed between the coherent and incoherent conditions (see Table 1), causing a waveform difference between the two conditions that could spill over into the ERPs of CW. In the two adjacent 50-ms time windows in the 150–250 ms interval, the incoherent conditions were more positive than the coherent conditions at anterior region (including Fz), resulting in an interaction of coherence with electrode or with region in each time window ($F(2,46)=18.57$–24.45, $ps<0.0005$, $MSE=1.14$–1.19). After that, there were no reliable differences across conditions until 450 ms after CW onset. In the two adjacent 50-ms time windows in the 450–550 ms interval, the ambiguous conditions were more positive than the unambiguous conditions at Fz, resulting in an interaction of referential ambiguity with electrode in each time window (both $F(2, 46)>5.65$, both $p<0.05$, $MSE=1.02$ or 0.83).

In nine adjacent 50-ms time windows in the 550–1000 ms interval, the incoherent conditions were more positive than the coherent conditions either at broad scalp sites, reflected in a main effect of coherence at the midline and lateral sites ($F(1,23)=6.85$–14.38, $ps<0.05$, $MSE=4.51$–8.29), or at centro-posterior regions (including Cz and Pz), reflected in an interaction of coherence with electrode or with region ($F(2,46)=7.08$–40.13, $ps<0.05$, $MSE=0.81$–1.91). In addition, in eight adjacent 50-ms time windows in the 600–1000 ms interval, the ambiguous conditions were more positive than the unambiguous conditions either at broad scalp sites, reflected in a main effect of referential ambiguity at the midline and/or lateral sites ($F(1,23)=4.30$–7.61, $ps<0.05$, $MSE=3.67$–6.41), or at centro-posterior regions (including Cz and Pz), reflected in an interaction of referential ambiguity with electrode or with region ($F(2,46)=5.63$–10.56, $ps<0.05$, $MSE=0.37$–1.18).

Three time windows were chosen on the basis of the results of these 50-ms interval analyses and for comparison with the results of Experiment 1: (a) 150–250 ms after critical word onset for early positivity effects to incoherence of the NP; (b) 350–550 ms for comparison with the results of Experiment 1, and (c) 550–1000 ms for late positivity effects to incoherence (this time window would also cover the time range of the late positivity effects to referential ambiguity). The results of the overall ANOVAs for the three time windows are shown in Table 3.

2.2.2.1.1. *The 150–250-ms time window*. At the midline sites, there was an effect of coherence and an interaction of coherence with electrode. Separate analyses found a larger positivity for the incoherent compared to the coherent nouns at Fz and Cz ($F(1,23)=26.51$ and 10.50, both $p<0.05$, $MSE=1.29$ and 0.85, respectively). In addition, there was an effect of

Table 3 – Experiment 2 overall analyses of variance for three time windows (in milliseconds) for midline and lateral electrodes ($N=24$).

Source	*dfs*	150–250			350–550			550–1000		
		F	*p*	*MSE*	F	*p*	*MSE*	F	*p*	*MSE*
Midline electrodes										
COH	1,23	11.08	0.003	4.60	<1			22.88	<.0005	2.50
REF	1,23	6.40	0.019	1.26	<1			6.97	0.015	2.89
COH × electrode	2,46	28.66	<.0005	0.87	<1			33.29	<.0005	0.54
REF × electrode	2,46	2.57	0.110	0.51	6.85	0.007	0.51	2.09	0.156	0.35
COH × REF	1,23	1.55	0.225	2.63	2.88	0.103	1.20	<1		
COH × REF × electrode	2,46	<1			2.10	0.154	0.45	<1		
Lateral electrodes										
COH	1,23	2.42	0.133	4.51	<1			21.00	<.0005	2.60
REF	1,23	2.22	0.150	1.12	<1			6.98	0.015	2.83
COH × region	2,46	28.19	<.0005	0.89	<1			26.67	<.0005	0.55
COH × hemisphere	1,23	2.93	0.100	0.42	<1			<1		
COH × region × hemisphere	2,46	<1			13.45	<.0005	0.05	<1		
REF × region	2,46	4.41	0.034	0.34	<1			4.00	0.045	0.23
REF × hemisphere	1,23	5.43	0.029	0.46	4.04	0.056	0.34	1.11	0.302	0.27
REF × region × hemisphere	2,46	<1			<1			<1		
COH × REF	1,23	<1			3.21	0.086	1.16	<1		
COH × REF × region	2,46	<1			1.12	0.305	0.46	<1		
COH × REF × hemisphere	1,23	1.51	0.232	0.33	1.45	0.242	0.27	<1		
COH × REF × region × hemisphere	2,46	<1			<1			<1		

COH = coherence; REF = referential ambiguity.

referential ambiguity, with a larger positivity for the ambiguous compared to the unambiguous nouns. For the lateral sites, there was an interaction of coherence with region, which was due to a larger anterior positivity for the incoherent compared to the coherent nouns ($F(1,23)=13.54$, $p=0.003$, $MSE=0.68$). There was also an interaction of referential ambiguity with region, which was due to a larger anterior positivity for the ambiguous compared to the unambiguous nouns ($F(1,23)=8.72$, $p=0.021$, $MSE=0.11$). In addition, there was an interaction of referential ambiguity with hemisphere, which was due to a larger positivity for the ambiguous compared to the unambiguous nouns at right hemisphere ($F(1,23)=8.05$, $p=0.018$, $MSE=0.10$). The early positivity, which was also observed in Experiment 1 (though only for incoherence), may reflect more difficult lexical processing or more attention for the incoherent or referentially ambiguous nouns compared to the coherent or unambiguous nouns.

2.2.2.1.2. *The 350–550-ms time window*. At the midline sites, although there was an interaction of referential ambiguity with electrode, separate analyses did not show any significant referential ambiguity effects ($F<1$ or $p>0.10$).

At the lateral sites, although there was an interaction among coherence, region, and hemisphere, a marginally significant interaction of referential ambiguity with hemisphere, and a marginally significant interaction of referential ambiguity with coherence, separate analyses did not show any significant coherence or referential ambiguity effects ($F<1$ or $p>0.17$).

2.2.2.1.3. *The 550–1000-ms time window*. At the midline sites, there was an effect of coherence and an interaction of coherence with electrode, which was due to a larger positivity (P600) at Cz and Pz for the incoherent compared to the coherent nouns ($F(1,23)=23.84$ and 48.61, both $p<0.0005$, $MSE=0.51$ and 0.57, respectively). In addition, there was an effect of referential ambiguity, with a larger positivity for the ambiguous compared to the unambiguous nouns.

At the lateral sites, there was an effect of coherence and an interaction of coherence with region. Separate analyses found a larger centro-posterior positivity (P600) for the incoherent compared to the coherent nouns ($F(1,23)=27.17$ and 42.43, respectively, both $p<0.0005$, both $MSE=0.25$). In addition, there was an effect of referential ambiguity and an interaction of referential ambiguity with region. Separate analyses found a larger centro-posterior positivity (P600) for the ambiguous compared to the unambiguous nouns ($F(1,23)=7.70$ and 9.16, both $p<0.05$, $MSE=0.28$ and 0.26, respectively).

2.2.2.2. *No-positivity-to-ambiguity group*. As in Nieuwland and van Berkum (2008a), our predictions depend upon the presence of the Nref effect in the AMBIGUOUS condition. In addition, participants showing a late positivity to referentially problematic expressions due to referential ambiguity or syntactic violation have been reported not to show an Nref to referential ambiguity or retrieval interference during the building of determiner-antecedent dependency (Martin et al., 2014; Nieuwland and van Berkum, 2008a). Therefore, like Martin et al. (2014) and Nieuwland and van Berkum (2008a), we adopted the analysis strategy of analyzing the subset of participants who did not exhibit a late positivity effect. Even in our Experiment 1, it is plausible that only a subset of our participants did not respond with a late positivity when confronted with a referential ambiguity, resulting in the Nref effect for the AMBIGUOUS condition being not very sustained for the analyses using the complete dataset, as we mentioned in Section 2.1.2. Therefore, for both experiments, we computed the mean difference of AMBIGUOUS minus CONTROL across all three midline electrodes and six lateral regions in the 550–1000-ms time window for each participant[3]. If the mean value in the AMBIGUOUS condition was numerically more positive than the mean value in the CONTROL condition, we classified the participant as producing a positivity response during that time window.

We found that 7/16 (44%) of the participants of Experiment 1 and 14/24 (58%) of the participants of Experiment 2 exhibited a positivity response to the AMBIGUOUS condition in the 550–1000-ms time window. The remaining 19 participants (9 from Experiment 1 and 10 from Experiment 2) exhibited a numerically more negative response to the AMBIGUOUS condition compared to the CONTROL condition. Analyzing only the 10 no-positivity participants from Experiment 2 would leave us insufficient power to test our predictions. Thus, we opted to combine the no-positivity participants from the two experiments and to treat experiment as a between-subject factor in the ANOVAs. These 19 no-positivity participants were classified as the no-positivity-to-ambiguity group, and the remaining 21 participants as the positivity-to-ambiguity group. The two groups did not differ in male/female ratio (10/9 and 10/11, respectively) and behavioral accuracy (93.72% and 93.00%, respectively; $t(38)=0.34$, $p=0.734$). The age difference between the two groups, although significant, was very small (20.6 and 22.0 years, respectively; $t(38)=2.44$, $p=0.020$). Below we report the results of both groups, although the no-positivity-to-ambiguity group is most informative to test the Blocking/Repair and the Independence hypotheses.

The results of the no-positivity-to-ambiguity group ($N=19$) are summarized in Fig. 3.

As shown in Fig. 3b and c, there was a sustained (400–1000 ms) negativity (Nref) effect with a broad distribution for the AMBIGUOUS condition. In addition, a typical P600 effect occurred for both the INCOHERENT and DOUBLE conditions. Most importantly, the effects for the DOUBLE condition were very similar to those for the INCOHERENT condition, rather than to those for the AMBIGUOUS condition, especially over Fz and F6 in the 400–550-ms interval and posterior regions (including Pz) in 400–1000-ms interval (all compared to CONTROL). Thus, it appears that the Nref effects were absent for the DOUBLE vs. INCOHERENT comparison.

These observations were statistically verified by ANOVAs with the factors referential ambiguity, coherence, electrode (midline) or hemisphere and region (lateral), and experiment (1/2), performed for each selected time window. We treated experiment as a factor in order to examine whether the factor

[3]The group split was based on the mean difference between the two conditions in the 550–1000-ms time window across all, rather than posterior or centro-posterior, scalp sites, as in Martin et al. (2014), because most (6/9 and 12/16 for Experiments 1 and 2, respectively) participants producing a late posterior positivity response showed a late positivity over broad instead of posterior or centro-posterior only scalp sites. An alternative group split based on the presence/absence of a centro-posterior positivity resulted in a very similar grouping and pattern of results.

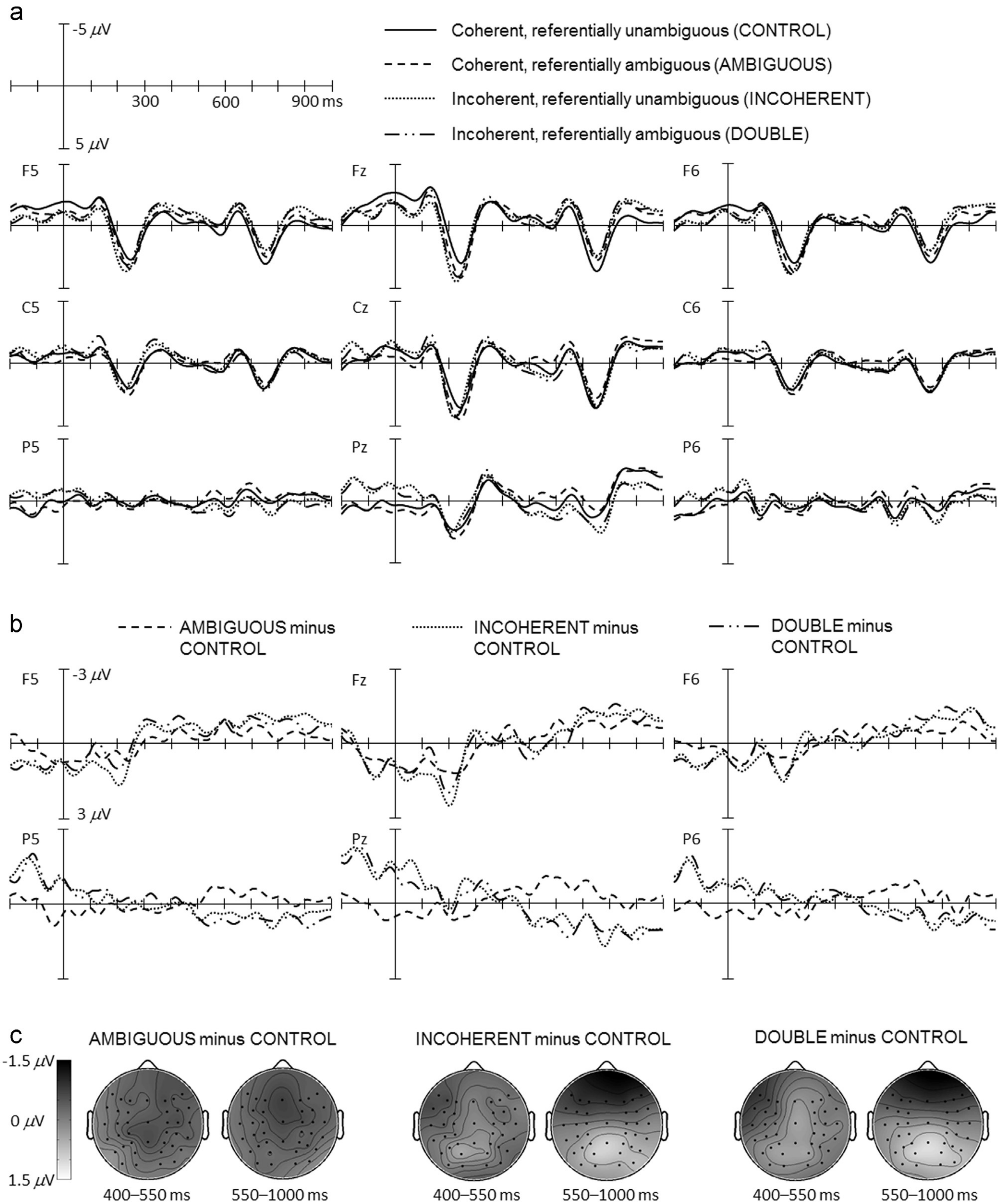

Fig. 3 – ERPs time locked to the onset of the critical words for all 19 participants who showed no positivity effects to the AMBIGUOUS condition in Experiments 1 and 2. (a) Grand average ERPs for all four critical conditions at nine scalp sites; (b) three difference waves at six scalp sites; (c) the scalp topographies of the three difference waves in two time windows.

experiment modulated the statistical mode of referential ambiguity–coherence interplay. In addition, in order to more objectively determine the time windows for the Nref effects of AMBIGUOUS vs. CONTROL, we performed analyses with the factors of experiment, condition (AMBIGUOUS vs. CONTROL), electrode or region and hemisphere in adjacent 50 ms time windows in the interval starting from 200 ms before and ending 1000 ms after CW (critical word) onset. The differences were considered reliable when they were significant in at least two adjacent time windows.

The 50-ms interval analyses found that the AMBIGUOUS condition was more positive than the CONTROL condition in the interval starting from 50 ms before and ending 300 ms after CW onset, reflected either in an effect of referential ambiguity at midline or lateral sites ($F(1,17)=4.56$–21.96, $ps<0.05$, $MSE=1.04$–3.74), or in an interaction of referential ambiguity with electrode ($F(2,34)=5.29$, $p=0.030$, $MSE=0.94$), with hemisphere ($F(1,17)=5.27$, $p=0.035$, $MSE=0.57$), or with experiment at midline sites ($F(1,17)=5.95$, $p=0.026$, $MSE=3.39$). These differences must be due, at least partly, to processing of stimuli presented prior to the CW. Note that the discourse context preceding CW differed between the AMBIGUOUS and CONTROL conditions (see (1) and (2) in Table 1), causing a waveform difference between the two conditions that could spill over into the ERPs of CW. After that, there were no reliable differences between the two conditions until 400 ms after CW onset. Crucially, the AMBIGUOUS condition was more negative than the CONTROL condition in each 50-ms interval (except for the 800–850 and 900–950 ms time windows) in the 400–1000 ms range. This negativity was reflected either in an effect of referential ambiguity ($F(1,17)=5.28$–18.61, $ps<0.05$, $MSE=1.14$–2.08), or in an interaction of referential ambiguity with electrode ($F(2,34)=11.90$, $p=0.001$, $MSE=0.42$), with region and hemisphere ($F(2,34)=6.70$, $p=0.009$, $MSE=0.09$), with experiment ($F(1,17)=4.28$–7.23, $ps<0.05$, except for $p=0.054$ for the 950–1000 ms window, $MSE=1.17$–3.52), with region and experiment ($F(2,34)=8.08$ and 9.31, $ps<0.005$, $MSE=0.32$ and 0.36), or with hemisphere and experiment ($F(1,17)=5.64$, $p=0.030$, $MSE=0.55$).

Although the negativity effects appeared to be sustained, we chose the 400–550 and 550–1000 ms interval as the earlier and later time windows for the negativity effects, respectively, in order to be close to the time windows chosen for the analyses using the complete dataset for both experiments. Below we report the results of statistical analyses performed on the mean amplitudes in the two selected time windows.

2.2.2.2.1. *The 400–550-ms time window*. At the midline sites, there was an interaction of coherence with experiment ($F(1,17)=7.12$, $p=0.016$, $MSE=4.64$) and an interaction among coherence, electrode, and experiment ($F(2,34)=4.33$, $p=0.046$, $MSE=1.40$). Separate analyses only found a main effect of coherence for Experiment 1, with a larger positivity for the incoherent compared to the coherent nouns ($F(1,8)=7.58$, $p=0.025$, $MSE=2.84$). Crucially, there was an interaction between referential ambiguity and coherence ($F(1,17)=5.43$, $p=0.032$, $MSE=1.03$). Separate analyses found a borderline significant larger negativity (Nref) for the ambiguous compared to the unambiguous nouns when the nouns were coherent ($F(1,18)=5.64$, uncorrected $p=0.029$, Bonferroni-corrected $p=0.058$, $MSE=0.37$). In contrast, there were no referential ambiguity effects when the nouns were incoherent ($F<1$). Note that the degrees of freedom differ for the omnibus ANOVA, which includes the factor Experiment, and the separate analyses that are licensed by the interactions that do not involve the factor Experiment.

At the lateral sites, there was an interaction of coherence with hemisphere ($F(1,17)=6.04$, $p=0.025$, $MSE=0.34$) or with experiment ($F(1,17)=8.88$, $p=0.008$, $MSE=3.71$), and an interaction among coherence, hemisphere, and experiment ($F(1,17)=6.84$, $p=0.018$, $MSE=0.34$). Separate analyses found an interaction of coherence with hemisphere for Experiment 1 ($F(1,8)=12.46$, $p=0.008$, $MSE=0.06$), which was due to a larger positivity for the incoherent compared to the coherent nouns at right hemisphere for Experiment 1 ($F(1,8)=13.42$, $p=0.006$, $MSE=0.27$). In addition, there was an interaction among coherence, region, and experiment ($F(2,34)=6.36$, $p=0.019$, $MSE=1.36$). Separate analyses only found a main effect of coherence for Experiment 1, with a larger positivity for the incoherent compared to the coherent nouns ($F(1,8)=5.81$, $p=0.043$, $MSE=0.91$). Although there was an interaction among coherence, region, and hemisphere ($F(2,34)=17.33$, $p<0.0005$, $MSE=0.07$), separate analyses did not find any significant effects of coherence ($F<1$ or $p>0.24$)[4].

Crucially, the global ANOVA at the lateral sites also revealed a borderline interaction of referential ambiguity with coherence ($F(1,17)=4.39$, $p=0.051$, $MSE=0.77$). Separate analyses found a borderline significant larger negativity (Nref) for the ambiguous compared to the unambiguous nouns when the nouns were coherent ($F(1,18)=4.60$, uncorrected $p=0.046$, Bonferroni-corrected $p=0.092$, $MSE=0.17$). However, there were no referential ambiguity effects when the nouns were incoherent ($F<1$).

In addition, planned comparisons did not find significant differences between the DOUBLE and INCOHERENT conditions at any midline electrodes or lateral regions for either experiment ($F<1$ or $p>0.86$), though a significant Nref effect for the AMBIGUOUS vs. CONTROL comparison at Cz, Pz, and lateral central region for Experiment 2 ($F(1,9)=9.64$–10.80, $ps<0.05$, $MSE=0.20$–0.69).

In order to examine whether referential ambiguity processing is blocked by the local incoherence, rather than the reverse, we tested whether the effects for DOUBLE were similar to those for INCOHERENT, rather than to those for AMBIGUOUS. Therefore, three separate analyses were performed with the factors of experiment, condition (non-control condition vs. CONTROL), electrode (midline) or region and hemisphere (lateral). At the midline sites, for both DOUBLE and INCOHERENT, there was a positivity for Experiment 1 ($F(1,8)=6.86$ and 5.21, $p=0.031$ and 0.052, $MSE=3.10$ and 3.59, respectively), reflected in an interaction of condition with experiment ($F(1,17)=6.92$ and 9.68, both $p<0.05$, $MSE=3.27$ and 2.89, respectively) and a reliably or marginally significant interaction among condition, electrode, and experiment ($F(2,34)=6.75$ and 3.73, $p=0.013$ and 0.062, $MSE=1.03$ and 1.26, respectively). For AMBIGUOUS, in contrast, there was only a main effect of condition, with the AMBIGUOUS

[4]On the basis of visual inspection (see Fig. 3b) and previous studies (e.g., Hagoort et al., 2003), an additional time window of 300–500 ms was chosen to investigate whether the incoherent phrases elicited a LAN (left anterior negativity) effect that has been observed for phrases containing a syntactic category violation (e.g., Hagoort et al., 2003). The global ANOVA at the lateral sites for this time window revealed a LAN effect for the incoherent phrases that was reliably significant for uncorrected analyses and marginally significant for corrected analyses ($F(1,18)=4.52$, uncorrected $p=0.048$, Bonferroni-corrected $p=0.096$, $MSE=0.71$), reflected in an interaction among coherence, region, and hemisphere ($F(2,34)=12.85$, $p=0.001$, $MSE=0.11$) and an interaction of coherence with hemisphere at anterior region ($F(1,18)=5.95$, $p=0.025$, $MSE=0.27$).

condition being more negative than the CONTROL condition ($F(1,17)=5.98$, $p=0.026$, $MSE=0.97$).

At the lateral sites, for both DOUBLE and INCOHERENT, there was an interaction of condition with experiment ($F(1,17)=8.39$ and 11.21, $p=0.010$ and 0.004, $MSE=2.60$ and 2.29, respectively), an interaction among condition, region, and hemisphere ($F(2,34)=11.42$ and 9.03, both $p<0.005$, $MSE=0.06$ and 0.05, respectively), an interaction among condition, region, and experiment ($F(2,34)=9.30$ and 5.89, $p=0.005$ and 0.023, $MSE=0.91$ and 1.08, respectively), and a reliably or marginally significant interaction among condition, hemisphere, and experiment ($F(1,17)=7.71$ and 3.84, $p=0.013$ and 0.067, $MSE=$ 0.38 and 0.53, respectively). Separate analyses found a reliably or marginally significant interaction of condition with region in Experiment 1 for both DOUBLE and INCOHERENT ($F(2,16)=8.77$ and 5.05, $p=0.015$ and 0.053, $MSE=0.53$ and 0.67, respectively), which was due to a centro-posterior or central positivity that was significant for uncorrected analyses (DOUBLE: central, $F(1,8)=6.06$, uncorrected $p=0.039$, $MSE=0.50$; posterior, $F(1,8)=$ 6.28, uncorrected $p=0.037$, $MSE=1.08$; INCOHERENT: central, $F(1,8)=5.44$, uncorrected $p=0.048$, $MSE=0.54$). In addition, separate analyses found a right positivity in Experiment 1 for DOUBLE ($F(1,8)=11.01$, $p=0.022$, $MSE=0.37$), reflected in an interaction of condition with hemisphere ($F(1,8)=8.46$, $p=0.020$, $MSE=0.13$). In contrast, for AMBIGUOUS, there was only a main effect of condition, with the AMBIGUOUS condition being more negative than the CONTROL condition ($F(1,17)=$ 4.66, $p=0.045$, $MSE=0.93$). These results suggest that the ERP effects for DOUBLE were highly similar to those for INCOHERENT, rather than to those for AMBIGUOUS.

2.2.2.2.2. *The 550–1000-ms time window*. At the midline sites, there was an effect of coherence ($F(1,17)=4.50$, $p=0.049$, $MSE=1.71$) and an interaction of coherence with electrode ($F(2,34)=50.07$, $p<0.0005$, $MSE=0.45$), which was due to a larger negativity at Fz ($F(1,18)=7.14$, $p=0.048$, $MSE=0.40$) and a larger positivity (P600) at Pz for the incoherent compared to the coherent nouns ($F(1,18)=35.92$, $p<0.0005$, $MSE=0.39$). In addition, there was an effect of referential ambiguity ($F(1,17)=6.10$, $p=0.024$, $MSE=0.60$). More interestingly, there was an interaction of referential ambiguity with coherence ($F(1,17)=5.66$, $p=0.029$, $MSE=1.31$). Separate analyses found a larger negativity (Nref) for the ambiguous compared to the unambiguous nouns when the nouns were coherent ($F(1,18)=25.20$, $p<0.0005$, $MSE=0.14$), but not when they were incoherent ($F<1$).

At the lateral sites, there was an interaction of coherence with region ($F(2,34)=57.52$, $p<0.0005$, $MSE=0.40$), which was due to a larger anterior negativity and a larger centro-posterior positivity (P600) for the incoherent compared to the coherent nouns (anterior, $F(1,18)=7.57$, $p=0.039$, $MSE=0.23$; central, $F(1,18)=$ 7.82, $p=0.036$, $MSE=0.16$; posterior, $F(1,18)=48.57$, $p<0.0005$, $MSE=0.12$). More interestingly, there was an interaction of referential ambiguity with coherence ($F(1,17)=7.10$, $p=0.016$, $MSE=0.82$). Separate analyses found a larger negativity (Nref) for the ambiguous compared to the unambiguous nouns when the nouns were coherent ($F(1,18)=16.18$, $p=0.002$, $MSE=0.07$). However, there was no difference between the ambiguous and unambiguous nouns when the nouns were incoherent ($F<1$).

In order to verify whether the effects for DOUBLE were similar to those for INCOHERENT, rather than to those for AMBIGUOUS, three separate analyses were performed, with the factors of experiment, condition (non-control condition vs. CONTROL), electrode (midline) or region and hemisphere (lateral). At the midline sites, there was an interaction of condition with electrode for both DOUBLE and INCOHERENT ($F(2,34)=29.66$ and 47.44, both $p<0.0005$, $MSE=0.47$ and 0.25, respectively), which was due to a negativity at Fz and a positivity at Pz for both DOUBLE and INCOHERENT (Fz, $F(1,18)=14.95$ and 9.31, $p=0.003$ and 0.021, $MSE=0.51$ and 0.86, respectively; Pz, $F(1,18)=14.39$ and 13.51, $p=0.003$ and 0.006, $MSE=0.69$ and 0.61, respectively). In contrast, for AMBIGUOUS, there was a main effect of condition, with the AMBIGUOUS condition being more negative than the CONTROL condition ($F(1,17)=24.23$, $p<0.0005$, $MSE=0.44$), without an interaction with electrode ($p>0.31$).

At the lateral sites, there was an interaction between condition and region for both DOUBLE and INCOHERENT ($F(2,34)=$ 26.16 and 49.27, both $p<0.0005$, $MSE=0.48$ and 0.25, respectively), which was due to an anterior negativity and a posterior positivity (P600) for both DOUBLE and INCOHERENT (anterior, $F(1,18)=11.51$ and 11.68, both $p=0.009$, $MSE=0.29$ and 0.39, respectively; posterior, $F(1,18)=22.54$ and 17.15, respectively, both $p<0.005$, both $MSE=0.21$). In contrast, for AMBIGUOUS, there was a main effect of condition, with the AMBIGUOUS condition being more negative than the CONTROL condition ($F(1,17)=15.40$, $p=0.001$, $MSE=0.44$), without an interaction with region ($p>0.25$). These results suggest that the ERP effects for DOUBLE were highly similar to those for INCOHERENT, rather than to those for AMBIGUOUS.

2.2.2.3. *Positivity-to-ambiguity group*. The results of the positivity-to-ambiguity group ($N=21$) are summarized in Fig. 4. For this group, a pattern emerges in which all three ambiguous and/or incoherent conditions elicited greater late positivity than the CONTROL condition. The two (400–550 and 550–1000 ms) time windows used for the no-positivity-to-ambiguity group were used for comparison with the results of that group. In the 400–550 ms time window, there were no reliable effects involving coherence ($F<1$ or $p>0.09$). For referential ambiguity, there was a negativity effect at both the midline and lateral sites in Experiment 1 ($F(1,6)=13.33$ and 14.05, $p=0.011$ and 0.020, $MSE=0.76$ and 0.09, respectively) and a marginally significant positivity effect at Fz in Experiment 2 ($F(1,13)=6.66$, $p=0.069$, $MSE=0.50$), reflected in an interaction among referential ambiguity, electrode, and experiment ($F(2,38)=6.67$, $p=0.009$, $MSE=0.67$), an interaction of referential ambiguity with electrode in Experiment 2 ($F(2,26)=4.51$, $p=0.030$, $MSE=0.27$), and an interaction of referential ambiguity with experiment at the lateral sites ($F(1,19)=$ 18.76, $p<0.0005$, $MSE=1.15$).

In the 550–1000 ms time window, both coherence and referential ambiguity exhibited a positivity effect at both the midline and lateral sites. There was a positivity effect to coherence at centro-posterior regions (including Cz and Pz) ($F(1,20)=19.40$–34.30, $ps<0.0005$, $MSE=0.27$–0.60), reflected in an interaction of coherence with electrode or with region ($F(2,38)=22.27$ and 17.98, both $p<0.0005$, $MSE=0.55$ and 0.58, respectively). In addition, there was an interaction of coherence with experiment at both the midline and lateral sites ($F(1,19)=5.23$ and 4.42, $p=0.034$ and 0.049, $MSE=2.01$ and 2.44, respectively), which was due to a positivity effect to coherence for Experiment 2 only

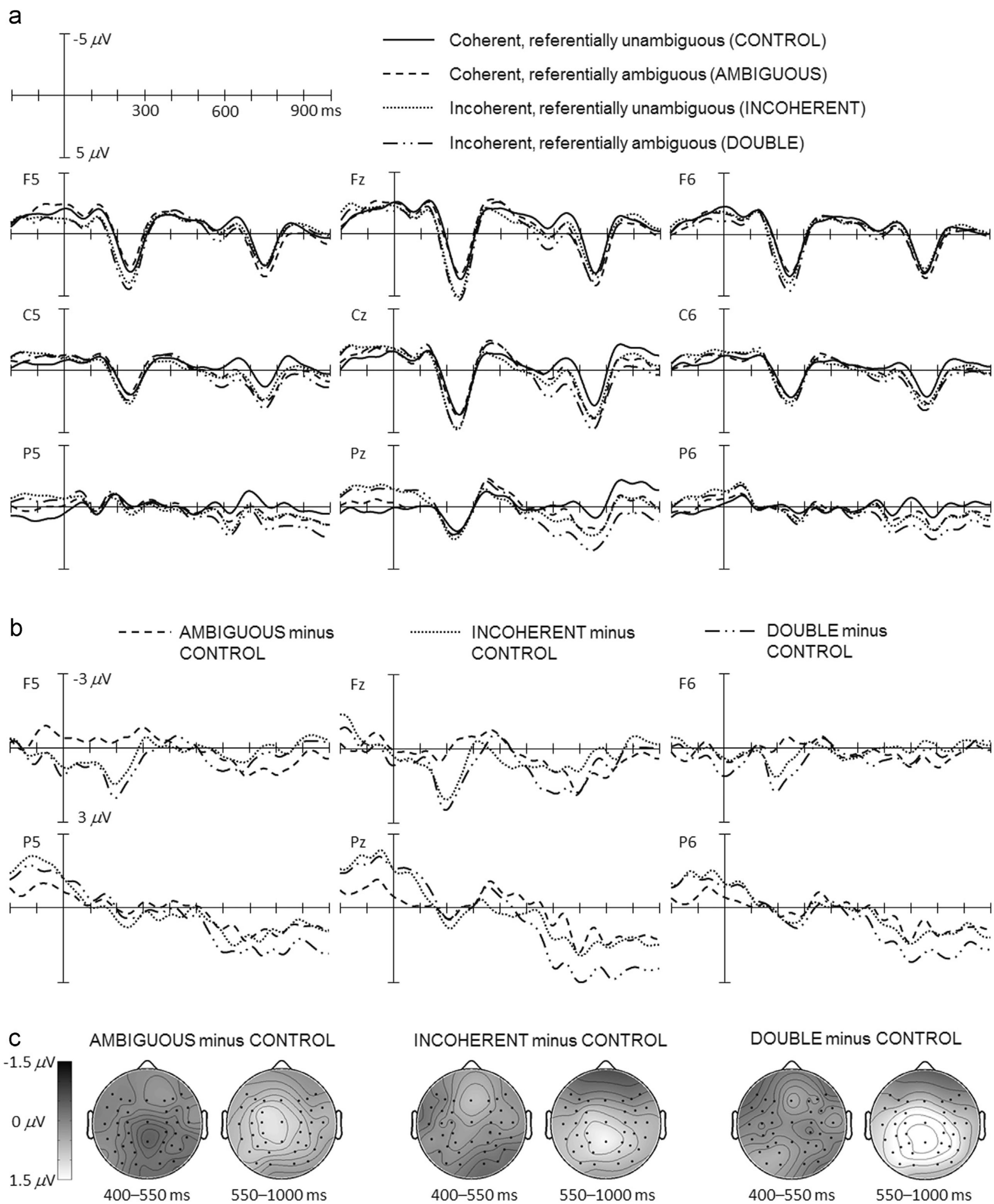

Fig. 4 – ERPs time locked to the onset of the critical words for all 21 participants who showed a positivity effect to the AMBIGUOUS condition in Experiments 1 and 2. (a) Grand average ERPs for all four critical conditions at nine scalp sites; (b) three difference waves at six scalp sites; (c) the scalp topographies of the three difference waves in two time windows.

($F(1,13)=21.97$ and 17.45, both $p<0.005$, $MSE=0.40$ and 0.24, respectively). For referential ambiguity, a positivity effect was evoked at both the midline and lateral sites ($F(1,19)=20.71$ and 18.71, both $p<0.0005$, $MSE=2.02$ and 1.90, respectively).

2.2.3. *Discussion*

While there was weak evidence for the Blocking Hypothesis in Experiment 1, such support was strengthened when the results of Experiment 2 were considered. Unlike in Experiment 1, a referential ambiguity elicited an Nref among less than half of participants in Experiment 2. When the Nref-producing (no-positivity-to-ambiguity) participants from the two experiments were considered together (with experiment as an additional factor), the results most clearly supported the Blocking Hypothesis: we observed a significant interaction between referential ambiguity and coherence, with an absence of an Nref effect for the DOUBLE condition and the ERP effects for DOUBLE and INCOHERENT being highly similar, both differing from those for AMBIGUOUS. We believe that the absence of the Nref effects were due to failures in phrase structure building, semantic processing, or both for the incoherent nouns in the DOUBLE (and INCOHERENT) condition.

The absence of the Nref effect due to a syntactic and/or semantic anomaly is highly similar to the "semantic blocking" observed by Nieuwland and van Berkum (2008a). Together, these results suggest that the referential processing reflected by the Nref is easily disrupted by local syntactic and/or semantic incoherence.

When considering the data from the no-positivity-to-ambiguity participants from both Experiments, the referential ambiguity (Nref) effect was rather sustained (from 400 to 1000 ms), being consistent with prior reports of Nref effects (e.g., Nieuwland et al., 2007; Nieuwland and van Berkum, 2006, 2008a; van Berkum et al., 1999, 2003, 2007; see Nieuwland and van Berkum, 2008b; van Berkum, 2004, for review). The fact that the Nref was already evident in the 400–550-ms time window is notable, given that the no-positivity-to-ambiguity group was formed according to the absence of positivity in the 550–1000-ms range. In addition, for the no-positivity-to-ambiguity group, the incoherence evoked a biphasic LAN-P600 response (the LAN occurred in the 300–500-ms time window, as reported in Footnote 4) that is typical for syntactic category anomalies (e.g., Hagoort et al., 2003).

As described in Section 2.2.2.1, the analyses of all 24 participants of Experiment 2 showed the local incoherence elicited a 550–1000-ms positivity effect that had a centro-posterior (including Cz and Pz) distribution, typical of P600 effects. Similar, though smaller and less broadly distributed, P600 effects were observed in Experiment 1. For referential ambiguity, however, the 350–550-ms negativity effect obtained in Experiment 1 was absent. Instead, a late positivity was evoked. The discrepancy between Experiments 2 and 1 is most likely due to the relatively low proportion of critical items and/or the involvement of a YES/NO comprehension task in Experiment 2. Both factors may increase the participants' attention to local incoherence and referential ambiguity, resulting in a relatively large P600 effect, which we will interpret below.

3. General discussion

In the present study, we manipulated both the coherence and the referential ambiguity of the subject NP in the third sentence of a three-sentence discourse. Like previous studies, we found a sustained (400–1000 ms) referential ambiguity (Nref) effect. The Nref effects were limited to a subset of our participants, as in previous studies (Martin et al., 2014; Nieuwland and van Berkum, 2008a). Thus, our primary findings – reported in Section 2.2.3 – combine the data from participants in both Experiments 1 and 2 who did not show P600 effects to coherent, referentially ambiguous nouns during the 550–1000 ms time window, with experiment treated as a factor.

As noted by some researchers (Martin et al., 2014; Nieuwland and van Berkum, 2008a), splitting participants into component-groups based on amplitude values could also split EEG background noise. In addition, the group split can render an effect (and related individual differences) trivial when the effect directly reflects the criterion used for the splitting of participants. Therefore, the group split in the present study can only be considered a more sensitive means for investigating how referential ambiguity processing is influenced by local coherence. We thus can only speculate that no-positivity-to-ambiguity participants most likely performed anaphoric inferencing when the nouns were coherent, resulting in an Nref effect to referential ambiguity for coherent NPs. In contrast, positivity-to-ambiguity participants might process the referentially ambiguous and/or incoherent nouns for purpose of an implicit judgment of acceptability or monitoring for processing errors, among others, resulting in a positivity effect to referential ambiguous and/or incoherent NPs (see Nieuwland and van Berkum, 2008a for a very detailed discussion of these possibilities).

In any case, the most important outcome of the component-group analysis is that Nref effects observed in the coherent condition were absent in both the earlier (400–550 ms) and later (550–1000 ms) time windows, when the NP was syntactically/semantically incoherent. This finding demonstrates that failures in phrase structure building, semantic processing, or both, due to the incoherence of the NP, block at least some aspects of referential interpretations, echoing Nieuwland and van Berkum (2008a).

One possibility is that the building of local syntactic and/or semantic coherence necessarily precedes referential processing, as assumed by the Blocking/Repair Hypothesis. Alternatively, local coherence and referential processes may occur in parallel, with referential processes proceeding over partially-complete phrases unless the phrase becomes syntactically or semantically anomalous. We look forward to future studies to address these possibilities. Either way, our finding provides evidence against a strong version of the Independence Hypothesis, in which referential interpretations still proceed normally, despite local syntactic and/or semantic incoherence of phrases.

In sum, our finding that some aspects of referential processing are blocked by local incoherence of phrases extends the findings of Nieuwland and van Berkum (2008a) to a different type of anomaly, which elicited a biphasic LAN-P600 instead of an N400-P600 response like that observed in Nieuwland and van Berkum (2008a). The fact that we found LAN-P600 effects may reflect the syntactic component of our incoherent nouns, which were both syntactically and semantically anomalous. If it is the case that the syntactic problem disrupted referential processing, our results would be consistent with all models that assume a functional primacy of syntactic processing over referential processing, such as versions of the extended argument-dependency model (Bornkessel and Schlesewsky, 2006; Bornkessel-schlesewsky and Schlesewsky, 2008, 2009). Future research may reveal whether other types of local syntactic or semantic incoherence, such as local morphosyntactic (e.g., gender) mismatch in case-marking languages like Dutch, French, and German, can block referential ambiguity processing (see Martín-Loeches et al., 2006 for a review of studies on how morphosyntactic processing interplays with semantic processing within a sentence). To get a further insight into the nature of the interplay between referential processing and the computation of local syntactic and semantic relationships, another interesting line for future research is probably to examine the mode of the interplay in bilingual people who process the input as second language.

Our finding also extends the findings of Nieuwland and van Berkum (2008a) to a different language. It is particularly interesting to find referential blocking in Chinese, given the prior finding that semantic integration of a verb and its object within a sentence is not blocked by a similar, local syntactic and semantic incoherence of the object NP when reading Chinese sentences (Zhang et al., 2010; Experiment 2). This suggests that discourse-level referential processing depends upon a coherent local phrase to a greater degree than sentence-level semantic processing, although discourse context information has been conjectured to play a very important role in sentence comprehension for Chinese (e.g., Xu, 1997).

4. Experimental procedures

4.1. *Experiment 1*

4.1.1. *Participants*

After giving informed consent, seventeen students from Peking University participated in Experiment 1. One participant (female) was not included because of a relatively low (89%) accuracy on the probe verification task (see below). Data from the remaining 16 (8 females; mean age 21 years, range: 18–25 years) were entered into the analyses. In this and the subsequent experiment, all participants were native speakers of Mandarin Chinese, were right-handed, had normal or corrected-to-normal vision, and had no known reading or neurological disorders. All were paid a nominal sum for their participation. This and the subsequent experiment were approved by the Academic Committee of the Department of Psychology, Peking University.

4.1.2. *Materials*

The critical materials were 160 sets of Chinese short discourses consisting of three sentences each (see Table 1 for examples). In each discourse, the first sentence introduced two human or object entities (e.g., "one brother and one sister" or "two brothers"), between which there was a clear contrast provided by the second sentence (e.g., "is very fat" and "is very thin"). The third sentence was the critical sentence, in which the subject NP was composed of the demonstrative pronoun *nage* ('*that*') and a noun in the coherent conditions (CONTROL and AMBIGUOUS), such as "*that brother*" in Table 1. For the incoherent conditions (INCOHERENT and DOUBLE), the degree adverb *hen* ('*very*') appeared between the pronoun and noun, such as "*that very brother*", resulting in an incoherence of the NP at the noun.

The subject nouns of the third sentences served as the critical words for ERPs. They had a single unique referent in the referentially unambiguous conditions (CONTROL and INCOHERENT), but had two equally eligible candidate referents in the referentially ambiguous conditions (AMBIGUOUS and DOUBLE). In addition, the critical words were lexically repeated for the same number of times (twice) in their preceding discourse context for each of the four conditions, to avoid any differential effects of lexical repetition across conditions.

The 160 sets of critical short discourses were assigned to four experimental lists using a Latin square procedure. For each list, the 160 critical items (40 for each condition) were pseudo-randomly mixed with 90 fully unproblematic filler short discourses, 50 of which had almost the same sentence construction as the critical items for the incoherent conditions, except for an adjective appeared between the degree adverb *hen* and the subject noun, to prevent participants from predicting the presence of a subsequent incoherence based on the reading of *hen*. The other 40 fillers had other sentence constructions. Therefore, there were 130 fully unproblematic and 120 problematic short discourses in total within each list.

4.1.3. *Procedure*

Participants were seated in a comfortable chair approximately 1 m from the computer screen in a dimly lit and sound-attenuated room. They read the short discourses sequentially, as each word (or sometimes a short phrase) appeared in the center of the screen. Each trial started with a central fixation cross presented for 800 ms, followed by a 500 ms blank screen. Each word or short phrase was presented for 400 ms, with an additional 100-ms inter-stimulus interval, during which the screen was blank. After the presentation of the last segment of the short discourse, there was an 800-ms blank, followed two fifth of the time by a probe word. Participants were asked to read the discourse carefully and to judge whether the probe had been present in the preceding short discourse ("YES probe") or not ("NO probe") by pressing buttons.

The probe verification task was used to ensure that participants pursued a continuous attentive reading. Using this task, previous ERP studies of sentence processing have observed reliable semantic violation (N400) effects and syntactic violation effects such as LAN (left anterior negativity) and P600 effects (e.g., Friederici et al., 1999; Gunter et al., 2000). Of the probes, 50% were YES probes. The content words at each word position of the discourses served equally often as the YES probes. The NO probes consisted of content words that were unrelated to the words in the discourses. The probe words of each category were evenly distributed across conditions (6/64 of probe words for the critical materials was the critical word). The probes remained on the screen until the participant had responded or for maximum 3 s. The next trial began after a 1-s interval.

Each participant received only one of the four experimental lists. The total 250 short discourses within each list were divided into five blocks of 50 trials. Prior to the experimental blocks, participants received a practice block of 12 trials. The experimental session lasted about 1 h.

4.1.4. *ERP recording and analysis*

The electroencephalogram (EEG) was recorded from 62 Ag/AgCl electrodes mounted in an elastic cap (Quik-Cap, NeuroScan Inc., Herndon, Virginia, USA). Recordings were referenced to the left mastoid but were re-referenced to linked mastoids offline. The horizontal electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye, and the vertical EOG was recorded from electrodes placed above and below the participants' left eye. Electrode impedances were kept below 5kΩ. The EEG and EOG were amplified with a band-pass from DC to 70 Hz and were recorded continuously with a digitization rate of 500 Hz. ERPs were additionally filtered off-line (10 Hz low pass, 24 dB/oct) for the plots only. All statistical analyses were performed on the data that were not low-pass filtered.

ERPs time-locked to the critical words were computed for each participant, condition, and electrode site. The epoch interval was 1200 ms, ranging from 200 ms before the onset of the critical word to 1000 ms after it. Note that the word preceding the critical word differed between the coherent and incoherent conditions (e.g., '*that*' and '*very*', respectively, in Table 1), which could result in spillover ERP activity into the ERPs of critical word and cause a baseline artifact in the ERPs for critical word when a baseline correction was performed (Osterhout et al., 2004; Steinhauer and Drury, 2012). Therefore, instead of using a baseline correction, we applied a 0.3 Hz high-pass filter procedure to the continuous EEG data, as suggested by Widmann et al. (2014). This procedure can remove the slow voltage shifts and avoid causing the baseline artifact at the same time, as in some previous studies (e.g., Wolff et al., 2008; see Widmann et al., 2014 for discussion).

All filtered epochs were evaluated individually for EOG or other artifacts. Epochs with amplitudes exceeding $\pm 70\,\mu V$ were excluded from the averages through artifact rejection. The overall rejection rate was 8.44%, equal for all four conditions (CONTROL, 7.81%; AMBIGUOUS, 10.94%; INCOHERENT, 7.81%; and DOUBLE, 7.19%).

Recall that the unique prediction of the Independence Hypothesis was that the DOUBLE condition might exhibit an Nref effect, relative to the INCOHERENT condition during the earliest window in which a reliable Nref effect was observed for the AMBIGUOUS condition. Therefore, besides omnibus ANOVAs, we also conducted planned comparisons (the least conservative statistical test) to test whether there was a difference between the DOUBLE and INCOHERENT conditions and between the AMBIGUOUS and CONTROL conditions.

The Greenhouse–Geisser correction was applied when evaluating effects with more than one degree of freedom in the numerator. In these cases, the original degrees of freedom and the corrected mean square error and probability levels are reported. In addition, to control for family-wise error rate, Bonferroni correction was used and unless otherwise stated, corrected p-values are reported whenever two or more simultaneous comparisons were made, except for the analyses in successive 50-ms time windows where differences were considered reliable when they were significant in at least two adjacent time windows, as mentioned in Section 2.1.2.

4.2. *Experiment 2*

4.2.1. *Participants*

After giving informed consent, thirty students from Peking University participated in Experiment 2. Six participants (3 females) were not included either because of a large number of slow voltage shift or electrocardiogram (ECG) artifacts, or because of unusually low (less than 72%) accuracy on the sentence verification task (see below). Data from the remaining 24 participants (12 females; mean age 22 years, range: 19–25 years) were entered into the analyses.

4.2.2. *Materials*

The critical materials were the same as those used in Experiment 1. The 160 sets of critical short discourses were assigned to four experimental lists using a Latin square procedure. In each list, the 160 critical items were mixed with 280 filler items, 80 of which had almost the same sentence construction as the critical items for the incoherent conditions (INCOHERENT and DOUBLE), except for an adjective appeared between the degree adverb *hen* ('very') and the subject noun, to prevent participants from predicting the presence of a subsequent incoherence based on the reading of *hen*. The remaining 200 items had other sentence constructions and consisted of 120 fully unproblematic short discourses and 80 short discourses with various anomalies involving grammatical aspect, number, classifiers, transitivity, or semantics. Therefore, there were 240 fully unproblematic and 200 problematic short discourses in total within each list.

4.2.3. *Procedure*

The procedure was identical to that used in Experiment 1, except the following. After the presentation of the last segment of the short discourse and the subsequent 800-ms blank, a verification sentence occurred one fifth of the time, which was either a paraphrase of a sentence within the discourse or a mis-paraphrase of it. Participants were asked to decide whether the verification sentence correctly expressed the content of the preceding discourse by pressing buttons. Each sentence within a single discourse was equally often referred to by the verification sentence, which correctly expressed the content of the preceding discourse 50% of the time. Moreover, each type of critical and filler discourse was equally often followed by a verification sentence. The verification sentences remained on the screen until the participant had responded or for a maximum of 3 s.

Each participant received only one of the four experimental lists. The total 440 short discourses within each list were divided into five blocks of 88 trials. Prior to the experimental blocks, participants received a practice block of 22 trials. The experimental session lasted about 2 h.

4.2.4. *ERP recording and analysis*

The method of ERP recording and data analyses was identical to that used in Experiment 1, except the following. First, the EEG and EOG were amplified with a band-pass of 0.05–100 Hz, with all statistical analyses being performed on the 0.3–70-Hz band-pass filtered data[5].

Second, the overall rejection rate was 8.57%, equal for all four conditions (CONTROL, 8.33%; AMBIGUOUS, 9.06%; INCOHERENT, 9.06%; and DOUBLE, 7.81%).

[5]In Experiment 1, DC recordings were used and thus no high-pass filter was applied during data acquisition, resulting in relatively high level of noise due to slow voltage shifts. Therefore, in Experiment 2, the 0.05-Hz high-pass filter was used during recording. The 100-Hz low-pass filter was used for one purpose that was unrelated to the present study (time-frequency analyses of 70–100 Hz EEG). In addition, a 70-Hz low-pass filter was applied during offline data processing in order to be consistent with Experiment 1, in which a 70-Hz low-pass filter was used during data recording.

Acknowledgments

This research was supported by grants from the National Natural Science Foundation of China (31371021, 31070898), the Natural Science Foundation of Beijing (7112069), and the Academy of Korean Studies Grant (AKS-2011-R36). The authors thank Ms Shuwei Xue for her assistance in running participants and preparing the figures.

REFERENCES

Altmann, G., Steedman, M., 1988. Interaction with context during human sentence processing. Cognition 30, 191–238.
Barton, S.B., Sanford, A.J., 1993. A case-study of anomaly detection: shallow semantic processing and cohesion establishment. Mem. Cogn 21, 477–487.
Boland, J.E., Blodgett, A., 2001. Understanding the constraints on syntactic generation: lexical bias and discourse congruency effects on eye movements. J. Mem. Lang. 45, 391–411.
Bornkessel, I., Schlesewsky, M., 2006. The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. Psychol. Rev. 113, 787–821.
Bornkessel-schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on 'semantic P600' effects in language comprehension. Brain Res. Rev. 59, 55–73.
Bornkessel-schlesewsky, I., Schlesewsky, M., 2009. The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. Lang. Ling. Compass 3, 19–58.
Cloitre, M., Bever, T.G., 1988. Linguistic anaphors, levels of representation, and discourse. Lang. Cogn. Processes 3, 293–322.
Ferretti, T.R., Kutas, M., McRae, K., 2007. Verb aspect and the activation of event knowledge. J. Exp. Psychol. Learn. Mem. Cogn 33, 182–196.
Fodor, J.D., Inoue, A., 1994. The diagnosis and cure of garden paths. J. Psycholinguist. Res 23, 407–434.
Friederici, A.D., Steinhuaer, K., Frish, S., 1999. Lexical integration: sequential effects of syntactic and semantic information. Mem. Cogn 27, 438–453.
Garnham, A., 2001. Mental Models and the Interpretation of Anaphora. Taylor and Francis, Philadelphia PA.
Garrod, S.C., Sanford, A.J., 1994. Resolving sentences in a discourse context: how discourse representation affects language understanding. In: Gernsbacher, M.A. (Ed.), Handbook of Psycholinguistics. Academic Press, San Diego, pp. 675–698.
Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. J. Cogn. Neurosci 12, 556–568.
Hagoort, P., 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. NeuroImage 20, S18–S29.
Hagoort, P., 2005. On Broca, brain, and binding: a new framework. Trends Cogn. Sci. 9, 416–423.
Hagoort, P., Wassenaar, M., Brown, C.M., 2003. Syntax-related ERP-effects in Dutch. Cogn. Brain Res. 16, 38–50.
Hanna, J.E., Brennan, S.E., 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. J. Mem. Lang. 57, 596–615.
Lü, S., Zhu, D., 1979. Yufa Xiuci Jianghua. (On Grammar and Rhetoric). Liaoning Education Press, Shenyang (Reprinted in 2002).
Martin, A.E., Nieuwland, M.S., Carreiras, M., 2012. Event-related brain potentials index cue-based retrieval interference during sentence comprehension. NeuroImage 59, 1859–1869.
Martin, A.E., Nieuwland, M.S., Carreiras, M., 2014. Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. Brain Lang. 135, 42–51.
Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., Sommer, W., 2006. Semantics prevalence over syntax during sentence processing: a brain potential study of noun–adjective agreement in Spanish. Brain Res. 1093, 178–189.
Nieuwland, M.S., Otten, M., van Berkum, J.J.A., 2007. Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. J. Cogn. Neurosci 19, 228–236.
Nieuwland, M.S., van Berkum, J.J.A., 2006. Individual difference and contextual bias in pronoun resolution: evidence from ERPs. Brain Res. 1118, 155–167.
Nieuwland, M.S., van Berkum, J.J.A., 2008a. The interplay between semantic and referential aspects of anaphoric noun phrase resolution: evidence from ERPs. Brain Lang. 106, 119–131.
Nieuwland, M.S., van Berkum, J.J.A., 2008b. The neurocognition of referential ambiguity in language comprehension. Lang. Ling. Compass 2, 603–630.
Osterhout, L., Bersick, M., McLaughlin, J., 1997. Brain potentials reflect violations of gender stereotypes. Mem. Cogn 25, 273–285.
Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., Inoue, K., 2004. Sentences in the brain: event-related potentials as real-time reflections of sentence comprehension and language learning. In: Carreiras, M., Clifton Jr., C. (Eds.), The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. Psychology Press, New York, NY, pp. 272–308.
Sanford, A.J., Garrod, S.C., 1989. What, when, and how? Questions of immediacy in anaphoric reference resolution. Lang. Cogn. Processes 4, SI235–SI262.
Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., Carlson, G.N., 1999. Achieving incremental semantic interpretation through contextual interpretation. Cognition 71, 109–147.
Steinhauer, K., Drury, J.E., 2012. On the early left-anterior negativity (ELAN) in syntax studies. Brain Lang. 120, 135–162.
Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.E., 1995. Integration of visual and linguistic information in spoken language comprehension. Science 268, 1632–1634.
Trueswell, J., Tanenhaus, M., 1991. Tense, temporal context and syntactic ambiguity resolution. Lang. Cogn. Processes 6, 303–338.
van Berkum, J.J.A., 2004. Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? In: Carreiras, M., Clifton Jr., C. (Eds.), The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. Psychology Press, New York, NY, pp. 229–270.
van Berkum, J.J.A., Brown, C.M., Hagoort, P., 1999. Early referential context effects in sentence processing: evidence from event-related brain potentials. J. Mem. Lang. 47, 147–182.
van Berkum, J.J.A., Brown, C.M., Hagoort, P., Zwitserlood, P., 2003. Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. Psychophysiology 40, 235–248.
van Berkum, J.J.A., Koornneef, A.W., Otten, M., Nieuwland, M.S., 2007. Establishing reference in language comprehension: an electrophysiological perspective. Brain Res. 1146, 158–171.
van de Meerendonk, N., Kolk, H.H.J., Chwilla, D.J., Vissers, C.T.W. M., 2009. Monitoring in language perception. Lang. Ling. Compass 3, 1211–1224.
Weber, A., Braun, B., Crocker, M.W., 2006. Finding referents in time: eye-tracking evidence for the role of contrastive accents. Lang. Speech 49, 367–392.
Widmann, A., Schröger, E., Maess, B., 2014. Digital filter design for electrophysiological data—a practical approach. J. Neurosci. Methods (in press).
Wolff, S., Schlesewsky, M., Hirotani, M., Bornkessel-Schlesewsky, I., 2008. The neural mechanisms of word order processing

revisited: electrophysiological evidence from Japanese. Brain Lang. 107, 133–157.

Xu, T., 1997. Yuyan lun. (On Language). Northeast Normal University Press, Changchun.

Zhang, Y., Yu, J., Boland, J.E., 2010. Semantics does not need a processing license from syntax in reading Chinese. J. Exp. Psychol. Learn. Mem. Cogn 36, 765–781.