The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs

Mante S. Nieuwland [a,b,c,*], Jos J.A. Van Berkum [a,d,e]

[a] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[b] Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA
[c] MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA
[d] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[e] F.C. Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

ABSTRACT

In this event-related brain potential (ERP) study, we examined how semantic and referential aspects of anaphoric noun phrase resolution interact during discourse comprehension. We used a full factorial design that crossed referential ambiguity with semantic incoherence. Ambiguous anaphors elicited a sustained negative shift (Nref effect), and incoherent anaphors elicited an N400 effect. Simultaneously ambiguous and incoherent anaphors elicited an ERP pattern resembling that of the incoherent anaphors. These results suggest that semantic incoherence can preclude readers from engaging in anaphoric inferencing. Furthermore, approximately half of our participants unexpectedly showed common late positive effects to the three types of problematic anaphors. We relate the latter finding to recent accounts of what the P600 might reflect, and to the role of individual differences therein.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Consider the meaning of the anaphoric noun phrase 'the necklace' in the following mini-story: "Britney Spears had several pieces of jewelry, including a golden necklace and a silver bracelet. One day, she was about to leave for a gala. She was admiring the *necklace* when suddenly her date called to cancel". To understand this passage, we have to combine the meaning of 'the necklace' with the preceding context to understand that Britney Spears thought some specific necklace was beautiful (semantic analysis), and simultaneously establish that 'the necklace' refers to the same piece of jewelry that was introduced in the first sentence (referential analysis). It has long been recognized that semantic and referential analysis co-determine our interpretation of anaphoric noun phrases (e.g., Garnham, 2001; Garrod & Sanford, 1994), and, more generally, that semantic and referential aspects of language comprehension are intricately interwoven at the functional level (e.g., Jackendoff, 2002). However, the exact mechanisms by which these two aspects of interpretation interact during language comprehension are as yet unknown.

Results from numerous psycholinguistic experiments have shown that anaphor resolution is greatly determined by semantic factors (e.g., Dell, McKoon, & Ratcliff, 1983; Garnham, Oakhill, & Cain, 1997; McKoon & Ratcliff, 1980; O'Brien, Raney, Albrecht & Rayner,1997; for reviews, see Almor, 1999; Ariel, 1990; Garnham, 2001; Myers & O'Brien, 1998; Garrod & Sanford, 1994; Gernsbacher, 1989; Sanford & Garrod, 1989). For example, readers experience difficulties in resolving anaphors that are more specific than their antecedent (e.g., "A bird would sometimes wander into the house. The robin was attracted by the larder."), but not vice versa (e.g., "A robin...The bird"; Garrod & Sanford, 1977; see also Garnham et al., 1997; Gordon & Hendrick, 1998). Furthermore, concepts central to real-world schemata are immediately available as referents for anaphoric noun phrases, even when they have not been explicitly mentioned (e.g., "Mary dressed the baby. The *clothes* were made of pink wool", Garrod & Sanford, 1981; (Walker & Yekovich, 1987).

In recent years, the behavioral literature on semantic aspects of anaphor resolution has been complemented with findings from event-related brain potentials (ERPs; e.g., see Burkhardt, 2006; Ditman et al., 2006; Nieuwland, Otten & Van Berkum, 2007a; Swaab, Camblin, & Gordon, 2004). For example, overly specific anaphors give rise to N400 effects (e.g., Swaab et al., 2004), reflecting the immediate difficulties with semantic processing these anaphors impose on our language comprehension system. Also, ERP results

* Corresponding author. Address: Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA. Fax: +1 617 627 3181.
E-mail address: mante@nmr.mgh.harvard.edu (M.S. Nieuwland).

have shown that verb-based implicit causality has an immediate impact on anaphor resolution, and that this impact can play out as a syntactic or semantic problem depending on which type of anaphor is involved (e.g., pronoun or proper name; Ledoux, Gordon, Roll, & Swaab, 2007; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). Finally, and of special importance to the current study, we have recently shown that discourse-level semantic manipulations can preclude referential ambiguity: whereas referentially ambiguous words (e.g., 'girl' in a two-girl context) elicit a frontally distributed, sustained negative shift compared to referentially unique words (Nref effect; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003), this referentially induced ERP effect does not occur if one of the referential candidates is an implausible antecedent (e.g., one of the girls died early in the story; Nieuwland, Otten, & Van Berkum, 2007a; see also Nieuwland & Van Berkum, 2006, for convergent evidence from ambiguous pronoun resolution).

The available evidence thus testifies to a delicate interplay between semantic and referential aspects of language comprehension. Moreover, results from a functional magnetic resonance imaging (fMRI) study (Nieuwland, Petersson, & Van Berkum, 2007b) seem to extend this notion to a neural systems level. In this study, referentially ambiguous pronouns selectively recruited medial prefrontal regions, suggesting that readers engaged in controlled inferencing (problem-solving) to evaluate the two antecedents in the discourse model and to make an anaphoric inference (Greene, McKoon, & Ratcliff, 1992). Intriguingly, we also observed a double dissociation for cortical networks involved in dealing with semantic anomalous and referentially ambiguous words. Brain regions activated by semantic anomaly were relatively deactivated by referential ambiguity, whereas regions activated by referential ambiguity were deactivated by semantic anomaly. These results may be taken to reflect the dynamic recruitment of semantic and episodic processing (e.g., Bookheimer, 2002; Wagner, Shannon, Kahn, & Buckner, 2005) to resolve semantically or referentially problematic situations.

In this study, we further examined the interplay and mutual dependence of semantic and referential analysis during anaphoric noun phrase resolution. More specifically, for reasons explained below, we investigated whether a problem with semantic analysis would preclude, aggravate or be inconsequential to referential analysis, and/or vice versa, this by orthogonally manipulating the semantic and referential status of an anaphoric noun phrase. As illustrated in Table 1 (see Appendix A for additional examples), we used stories in which the critical noun phrase was either semantically coherent (e.g., "She was admiring the *necklace*…") or incoherent (e.g., "She stepped into the *necklace*…"), and in which either of these noun phrases did or did not have a unique antecedent in the preceding discourse Thus, the critical noun phrase could be fully unproblematic, i.e. unambiguous and coherent, referentially ambiguous only, semantically incoherent only, or simultaneously ambiguous and incoherent. To keep track of semantic and referential analysis relatively selectively, with high temporal resolution we recorded ERPs as our participants read through these stories.

Of course, in spontaneously occurring language, the simultaneous occurrence of referential ambiguity and semantic incoherence is a relatively unlikely, and a rather serious event. However, this does not mean that the comprehender's brain response will thus be uninformative. Note, first of all, that referential ambiguity by itself is not at all uncommon in everyday language use (e.g., Auer, 1984; Trueswell & Tanenhaus, 2005), and the typical ERP response observed to this situation can be assumed to reflect the routine processing that comprehenders engage in to resolve the ambiguity (Nieuwland et al., 2007a; Van Berkum et al., 2007). Furthermore, we know that the ERP response to semantic violations indexes the very neurocognitive processes that are also involved in the routine semantic analysis of normal linguistic expressions (e.g., Brown, Hagoort, & Kutas, 2000; Kutas & Hillyard, 1984; Kutas & Van Petten, 1994). Finally, and as evidenced by the fruitful use of other 'double anomalies' in ERP research (e.g., Hagoort, 2003; Osterhout & Nicol, 1999), anomaly detection taps into the sensitivity of the comprehension system, and potential interactions between different types of anomalies can thus inform us about what analysis at one level does to analysis at another level. In our specific case, the observation that, for example, one conceptual processing stream A (e.g., semantic analysis) can proceed when another conceptual processing stream B (e.g., referential analysis) cannot deliver, would constitute strong evidence for the functional independence of process A relative to process B.

We should note that whereas anaphor resolution is conceptualized as a semantic task in some linguistic accounts (see Chierchia & McConnell-Ginet, 1990; Kamp & Reyle, 1993), the inferences needed to disambiguate reference are often defined as pragmatic inferences (see Levinson, 2000; Wilson & Sperber, 2004). Although the psychological reality of the distinction between semantic and pragmatics is under debate (e.g., Gibbs, 1984; Jackendoff, 2002), one could argue that our study also speaks to the issue of how the semantic meaning of a word in a particular sentence context can influence the computation of its pragmatic meaning.

Table 1
Example story (approximate English translation from Dutch, and original Dutch story)

	1-Referent introduction	2-Referent introduction
Semantically coherent continuation	*Semantically coherent + referentially unique control* Britney Spears had several pieces of jewelry, including a golden necklace and a silver bracelet. One day she was about to leave for a gala. She was admiring the necklace when suddenly her date called to cancel. Britney Spears had meerdere sieraden, waaronder een gouden ketting en een zilveren armband. Op een dag stond ze klaar om naar een gala te gaan. Ze bewonderde de ketting toen haar date opeens afbelde.	*Referentially ambiguous* 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 was admiring the necklace when suddenly her date called to cancel. Britney Spears had meerdere sieraden, waaronder een gouden ketting en ook een van zilveren. Op een dag stond ze klaar om naar een gala te gaan. Ze bewonderde de ketting toen haar date opeens afbelde.
Semantically incoherent continuation	*Semantically incoherent* Britney Spears had several pieces of jewelry, including a golden necklace and a silver bracelet. One day she was about to leave for a gala. She stepped into the necklace when suddenly her date called to cancel. Britney Spears had meerdere sieraden, waaronder een gouden ketting en een zilveren armband. Op een dag stond ze klaar om naar een gala te gaan. Ze stapte in de ketting toen haar date opeens afbelde.	*Semantically incoherent + Referentially ambiguous* 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 when suddenly her date called to cancel. Britney Spears had meerdere sieraden, waaronder een gouden ketting en ook een van zilveren. Op een dag stond ze klaar om naar een gala te gaan. Ze stapte in de ketting toen haar date opeens afbelde.

Critical noun phrases are underlined for expository purpose. Four different story versions were presented (the semantically coherent + referentially unique control condition, the referentially ambiguous condition, the semantically incoherent conditions, and the semantically incoherent + referentially ambiguous condition).

Our predictions for referential ambiguity and semantic incoherence separately (each compared to the control condition) were straightforward. We expected referentially ambiguous anaphors to elicit an Nref effect, the frontally-distributed, negative shift that is typically associated with referential ambiguity (Van Berkum et al., 1999; Van Berkum et al., 2003), and we predicted that semantically incoherent anaphors would elicit a standard N400 effect (Kutas & Hillyard, 1980). Our critical comparison, however, involved the combination of referential ambiguity and semantic incoherence. As far as we can tell, extant models of anaphor resolution, such as the Bonding and Resolution framework (e.g., Sanford & Garrod, 1989; see also Garnham, 2001 for a review), do not lead to strong predictions for this case.[1] However, one can easily distinguish several logical possibilities about how the language comprehension system would cope with these combined violations.

Firstly, semantic incoherence could preclude readers from engaging in elaborate anaphoric inferencing. This prediction follows from a "standard of coherence" perspective (e.g., Van den Broek, Risden, & Husebye-Hartmann, 1995): when anaphors pose a powerful semantic coherence break, readers may devote all of their attentional resources to re-establishing semantic coherence, leaving too little to attend to the referential ambiguity conveyed by those same anaphors. Under such circumstances, readers could fail to resolve the 'full' referential implications of an ambiguous anaphor (e.g., Klin, Weingartner, Guzmán, & Levine 2004; Klin, Guzman, Weingartner, & Ralano 2006; Levine, Guzmán, & Klin, 2000; see also Ferreira, Bailey, & Ferraro, 2002; Garrod & Sanford, 1999; Sanford & Sturt, 2002), leading to the prediction that the ambiguous/incoherent anaphors elicit an N400 effect but no Nref effect (i.e., under-additivity).

As a second possibility, the combined presence of the two types of violations could 'boost' the impact of either one of the anomalies (i.e., over-additivity; see Hagoort, 2003, for a 'syntactic boost' effect of syntactic violations on the N400 effect elicited by semantic anomalies). One way this might go is that referential ambiguity adds to the unexpectedness of a semantically incoherent anaphor (equivalent to the 'syntactic boost'; Hagoort, 2003), causing the combined violations to elicit a larger N400 effect than single semantic anomalies. Alternatively, referential ambiguity resolution involves making an anaphoric inference to decide between the referential candidates (Nieuwland et al., 2007b), and this process may be more difficult when the antecedents are both implausible than when they are both plausible, as the ambiguity now has to be resolved with even fewer cues to which antecedent fits best. This could cause the combined violations to elicit a larger referential ambiguity ERP effect than referential ambiguity alone.[2]

Finally, it could be the case that referential ambiguity resolution is relatively independent of semantic analysis (both in terms of informational encapsulation and processing resources) as long as semantic analysis does not readily provide clues to resolve the ambiguity. In such a scenario, referential ambiguity resolution pro-

ceeds undisturbed in the face of semantic violations, causing the combined violations to elicit an Nref effect in addition to an N400 effect (i.e., additivity).

2. Methods

2.1. Subjects

After giving informed consent, 39 right-handed college students (8 males, mean age 21.1) participated in this study for course credit. All subjects were native speakers of Dutch, and without any neurological impairment.

2.2. Materials

Subjects were presented 120 stories consisting of two naturally spoken Dutch sentences followed by one visually presented sentence. In each story, the first sentence introduced the story protagonist and two secondary entities (persons or objects). Protagonists were denoted with proper names or NPs (e.g., "Britney Spears", "the cleaning lady"), whereas the secondary entities were always denoted with NPs (e.g., "a golden necklace"). In the 1-Referent condition, the secondary entities were denoted with different NPs (e.g., "a golden necklace and a silver bracelet"). In the 2-Referent condition, they were denoted by identical NPs that were preceded by distinct predicates, and the substitutions 'one' or 'another' were used to circumvent any differential effect of lexical repetition of the critical NPs across conditions (e.g., "a golden necklace and a silver one"). The second sentence was used to foreground the protagonist. In the final sentence (which was presented word-by-word on a computer screen using rapid serial visual presentation, see below), the critical NPs denoting one of the secondary entities were either semantically incoherent (e.g., "She stepped into the *necklace*...") or coherent (e.g., "She was admiring the *necklace*..."). In most stories, this semantic anomaly involved a strong lexical-semantic anomaly (e.g., "It was broadcasted live on the secretary...", "Jim made room in the colonel..."), but in some stories it involved a discourse-level semantic incoherency (e.g., "Lucy called up the burglar" after she just caught the burglar in her house; see Van Berkum et al., 2003, for evidence that this distinction doesn't matter to the N400 effect). In most of the stories, the coherent and incoherent final sentences differed only in one adverb (e.g., "Jim made room in/for the colonel...") in order to minimize pre-CW lexical differences. Also note that the critical words were inanimate (e.g., 'necklace' in the example story) only in about $\frac{1}{4}$ of all stories, and animate in the rest. Whether the secondary entity had been mentioned first or second in the first sentence was balanced across the set of stories. In all final sentences, at least four words following the critical word (CW) were kept identical across the coherent and incoherent version.

The two versions of each story introduction (1-Referent/2-Referent) were recorded by the same female speaker with a normal speaking rate and intonation. We used a 2 (Referential context: 1-Referent, 2-Referent introduction) × 2 (Coherence: coherent continuation, incoherent continuation) full factorial design (see Table 1; see Appendix A additional examples). This design rendered the CW in the final sentence semantically coherent but referentially ambiguous (referential ambiguity condition), semantically incoherent but referentially non-ambiguous (semantic incoherence condition), both semantically incoherent and referentially ambiguous (ambiguity/incoherence condition), or semantically coherent and referentially non-ambiguous (control condition).

---

[1] The Bonding & Resolution framework distinguishes between the initial localization/identification of an antecedent (bonding) and the ultimate full commitment to one particular interpretation (resolution). Here, we refrained from deriving specific predictions within this framework because we believe there is no simple one-to-one mapping between, on the one hand, semantic processes as indexed by the N400 ERP and referential processes as indexed by the Nref effect, and, on the other hand, bonding and resolution processes. For example, referential ambiguity can be thought to incur a bonding problem if only unique bonds between anaphor and antecedent are allowed, but a resolution problem if simultaneous bonds are allowed. Similarly, it is unclear whether a semantically incoherent anaphor would constitute a problem with either only binding or resolution, or perhaps with both.

[2] One could also argue for a combination of our last two possibilities, wherein both semantic processing and referential ambiguity resolution are more difficult in each other's presence so that *both* the expected ERP effects are 'boosted'.

2.3. Procedure

Participants were seated in front of two loudspeakers, and were informed that they would be listening to short stories of which the final sentence was always visually presented. They were instructed to listen and read for comprehension and minimize movement. No additional task demands were imposed.

Each trial was separated from the next by a 5-s silence and was preceded by a short warning tone. At 500 ms before offset of the spoken part, the visual presentation of a fixation mark (for the duration of 1 s) heralded the start of the visual presentation of the final sentence, which began 1 s after offset of the spoken part. To parallel natural reading times (e.g., Legge, Ahn, Klitz, & Luebker, 1997), all words (except from the word that directly preceded the critical pronouns onwards) were presented using a variable serial visual presentation procedure, wherein word duration in milliseconds was computed as ((number of letters * 30) + 190), with a maximum of 430 ms. However, to avoid spurious ERP effects due to differences in word length, we switched to non-variable presentation from the word that preceded the critical pronoun onwards (using a word duration of 300 ms, with an exception for final words, which were presented 800 ms). Importantly, participants did not notice this alternation between variable and fixed word duration presentation within sentences (see also Nieuwland & Van Berkum, 2006; Otten & Van Berkum, 2007). All inter-word-intervals were 200 ms.

Four trial lists were used. For the first list, 30 stories of each condition were pseudorandomly mixed with 60 filler stories (these were based on materials from Nieuwland & Van Berkum, 2005, though shortened and adapted to match the current visual presentation of the final sentence) such that neither condition occurred more than two times consecutively, while trials of each type were matched on average list position. The other three lists were derived from the first by rotating the different trial types. The total of 180 stories was divided in six blocks, separated by a pause. Total time-on-task was approximately 70 min.

2.4. EEG recording and pre-processing

The EEG was recorded from 30 standard scalp locations (Fz, Cz, Pz, Oz, Fp1/2, F3/4, F7/8, F9/10, FC1/2, FC5/6, FT9/10, C3/4, T7/8, CP1/2, CP5/6, P3/4, and P7/8) and amplified (band-pass filtered at 0.03–100 Hz) using a BrainAmps amplifier (BrainProducts, München), digitized at 500 Hz and re-referenced to the mean of left and right mastoids. Ocular and muscular artifacts were corrected by means of a procedure based on Independent Component Analysis (e.g., Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). Then, epochs that ranged from −500 to +1600 ms relative to critical word onset were extracted and normalized (by subtraction) to a 150 ms pre-onset baseline. Subsequently, segments with potentials exceeding ±75 μV were rejected, and the remainder was screened for drift artifacts. There was an average segment loss of 11% across all participants, but none of the participants met our exclusion criteria regarding rejection rates of artifactual epochs (a total rejection rate exceeding 40% of the trials).

3. Results

3.1. All subjects

As shown in Fig. 1 (upper graphs), referentially ambiguous anaphors (dashed green lines) elicited a sustained negative deflection at anterior electrodes compared to the control anaphors (solid black lines), consistent with an Nref effect, the known ERP signature of referential ambiguity (e.g., Nieuwland et al., 2007a; Van

Berkum et al., 1999; Van Berkum et al., 2003). Semantically incoherent anaphors (dotted red lines) elicited a slightly right-lateralized N400 effect, followed by a large positive deflection (or late positive component/LPC) at posterior electrodes. Crucially, the ERPs elicited by the ambiguous/incoherent anaphors (dashed/dotted blue lines) closely resembled the ERPs elicited by the incoherent anaphors: a slightly right-lateralized N400 effect, also followed by a large posterior LPC. The corresponding difference waveforms at electrode locations Fz and Pz, and their scalp distributions for the 300–500 and 500–1600 ms time windows are displayed in Fig. 1 (lower graphs). These time windows were also chosen for statistical analyses because it allowed us to examine N400 modulations within the standard N400 analysis window, while preserving maximal discriminatory power for the more slow and sustained ERP response to referential ambiguity. Also, because of the known distributions of ERP effects associated with referential ambiguity (anterior maximum of Nref effects) and semantic incoherence (posterior maximum of N400 effects), we performed statistical tests using only anterior (F3/4, F7/8, F9/10, FC1/2, FC5/6, FP1/2, FT9/10, and Fz) or posterior (CP1/2, CP5/6, Oz, P3/4, P7/8, and Pz) channels in addition to statistical tests that involved all electrodes. All statistical tests as presented here involve pairwise comparisons of each condition compared to the unambiguous, coherent control condition, pooled across individual electrodes.

Statistical analysis (see Table 2) confirmed that relative to correct controls, referentially ambiguous anaphors elicited a reliable Nref effect in the 500- to 1600-ms latency range, but no reliable N400 effect. Both observations are consistent with earlier findings (e.g., Nieuwland & Van Berkum, 2006; Nieuwland et al., 2007a; Van Berkum et al., 1999; Van Berkum et al., 2003). Furthermore, semantically incoherent anaphors elicited a reliable N400 effect, as well as, thereafter, a reliable posterior LPC. As might be expected from Fig. 1, the critical ambiguous/incoherent anaphors elicited a similarly reliable posterior LPC. However, in contrast to what the difference waves in Fig. 1 suggest, statistical tests in the N400 time window failed to reveal a significant N400 effect to ambiguous/incoherent anaphors.

The latter two observations appear inconsistent with any of the predictions that we put forward in the introduction, as we predicted that referential ambiguity would elicit only an Nref effect, and that the combined anaphors would elicit a differential effect (N400 effect and/or Nref effect) compared to the fully coherent control condition. However, closer examination of the difference waves in Fig. 1 suggests that ambiguous/incoherent anaphors did in fact increase the amplitude of the N400, but to an extent that was largely counteracted by an early LPC effect in that condition. A subsequent examination of single-subject effects elicited by each of our three problematic anaphors (ambiguity, incoherence, and ambiguity/incoherence) also revealed large individual differences in whether problematic anaphors elicited an LPC effect. These two observations led us to more closely examine the potential impact of temporally overlapping LPC effects on our results by means of an additional split-group analysis.

3.2. LPC-to-ambiguity group vs. no-LPC-to-ambiguity group

Our starting point was the ERP effect to referentially ambiguous anaphors. We had obtained a very small but reliable Nref effect when averaging across all subjects in that condition. However, visual inspection of the individual subject waveforms revealed that ambiguous anaphors elicited only a small average effect because approximately half of our subjects had generated a large LPC, instead of an Nref effect, to these anaphors. This suggests that these subjects might be engaged in comprehension processes other than referential ambiguity resolution, an issue that we will return to in our discussion. More importantly for current purposes, we
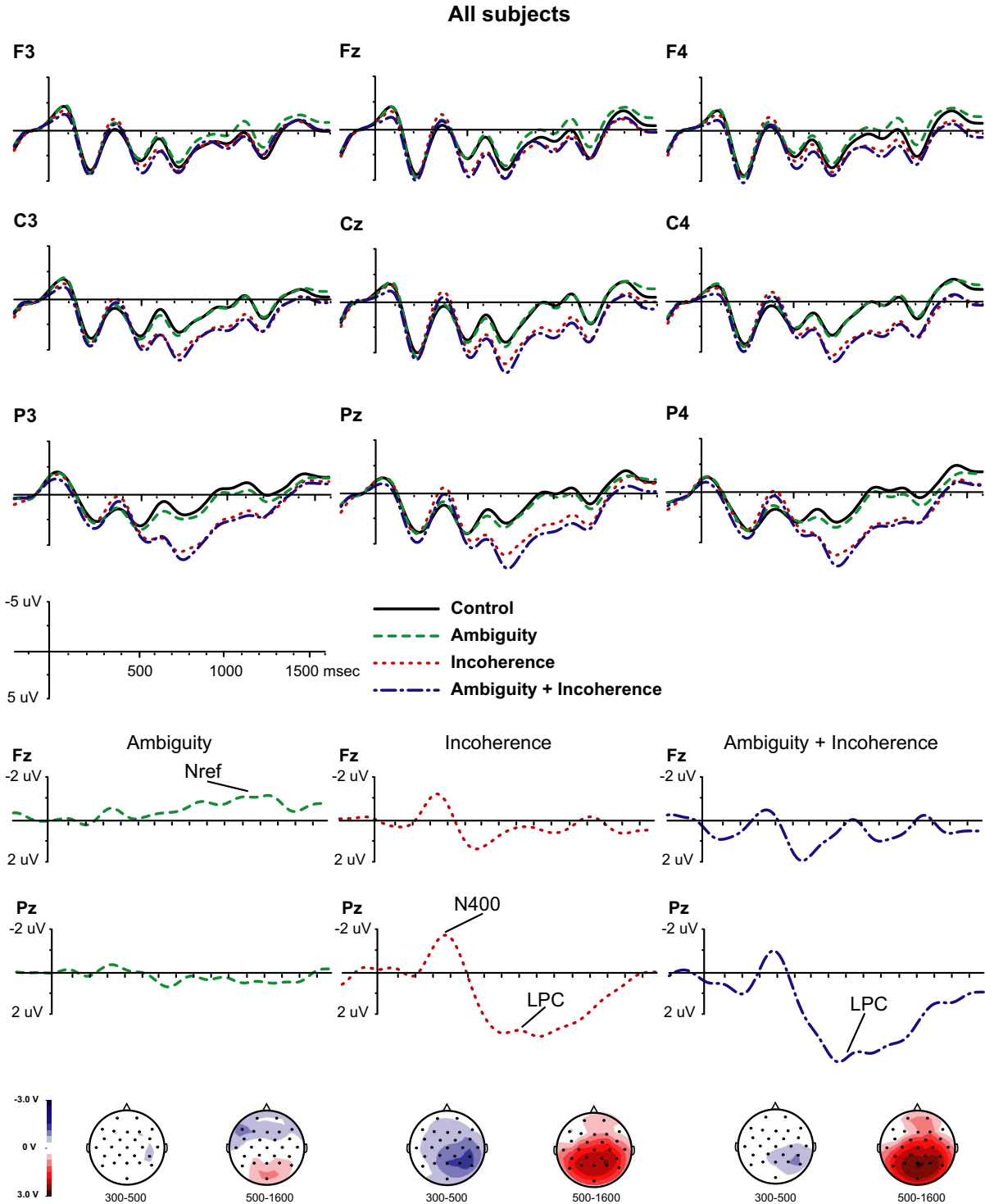
Fig. 1. All subjects. (Upper graph) ERPs to referentially ambiguous anaphors, semantically incoherent anaphors, ambiguous/incoherent anaphors and the control anaphors at a subset of electrode positions. (Lower graph) Difference waves of ambiguity, incoherence and ambiguity/incoherence (compared to the control condition) at Fz and Pz, and corresponding scalp distributions for the two time windows that were used for statistical analyses. In this and all following figures, negativity is plotted upwards, waveforms are filtered (5 Hz high cut-off, 48 dB/oct) for presentation purpose only, and we added the relevant ERP effect labels in the difference waves graphs.

reasoned that to the extent that subjects were not trying to resolve the referential ambiguity in the way we had consistently observed in earlier studies, we would not be in a position to examine the additional impact of a semantic anomaly on reference resolution. Simply put, without a referential negativity indexing the listener's natural attempts to resolve reference in this critical benchmark condition, we cannot see whether or not these reference resolution

processes would change in the face of an additional semantic problem.

To circumvent this problem, we divided our subjects into a group that showed an LPC response to referentially ambiguous anaphors and a group that did not show an LPC effect. Using the sign of the mean difference between the referential ambiguity and control condition (referential ambiguity minus control) at

Table 2
F, MSE, and p values for main effects of ambiguity, incoherence, and their combination in two consecutive time windows (computed across all electrodes and for anterior and posterior electrodes separately)

	Electrodes	300–500			500–1600		
		F	MSE	p	F	MSE	p
Ambiguity	All	.747	28.064	.393	.471	44.431	.497
	Anterior	.989	15.539	.326	4.727	20.160	.036[*]
	Posterior	.207	15.869	.651	1.112	21.691	.298
Incoherence	All	6.293	39.001	.017[*]	14.249	47.226	.001[**]
	Anterior	2.571	19.799	.117	1.774	26.456	.191
	Posterior	8.337	21.199	.006[**]	34.272	21.620	.000[***]
Combined	All	.264	53.742	.611	22.764	46.682	.000[***]
	Anterior	.000	25.258	.988	4.046	22.846	.051[^]
	Posterior	.967	26.459	.332	48.360	22.989	.000[***]

All statistical tests in this table and all following tables involve pairwise comparisons of each condition compared to the unambiguous, coherent control condition. For all F tests, numerator df = 1, denominator df = 38, $^\wedge p \leqslant .1$, $^* p \leqslant .05$, $^{**} p \leqslant .01$, $^{***} p \leqslant .001$.

posterior channels (CP1/2, CP5/6, Oz, P3/4, P7/8, and Pz) in the 500- to 1600-ms time window for each subject as a criterion, 19 subjects were classified as belonging to the LPC-to-ambiguity group, and the remaining 20 subjects were classified as the no-LPC-to-ambiguity group. The two groups were comparable in terms of male/female ratio (the groups consisted of 3 males and 4 males, respectively) and age (21.8 and 20.5 years respectively; t(37) = 1.46, p = .153), and did not differ in their performance on the Reading Span Test (68 and 73 respectively; t(37) = 1.26, p = .217; Reading Span performance was assessed following the main experiment, see Nieuwland & Van Berkum, 2006, for details on the procedure). Note that although the findings for the former group are interesting in their own right, the ERP effects displayed by the latter group should be most informative to the main research question addressed in this study.

ERP waveforms and difference waves for the no-LPC-to-ambiguity group are plotted in Fig. 2. In this critical group, ambiguous anaphors elicited a referentially induced ERP effect, whereas the incoherent anaphors and the ambiguous/incoherent anaphors each elicited an N400 effect followed by an LPC. ERP waveforms and difference waves for the LPC-to-ambiguity group are plotted in Fig. 3. In this group, anaphors from all three conditions elicited an LPC only.

As shown in Table 3, the group split resulted in significant condition[*] group interaction effects in the referential ambiguity condition (compared to the control condition) in both latency ranges. Ambiguous anaphors elicited a more negative ERP than control anaphors in the no-LPC-to-ambiguity group, but a more positive ERP in the LPC-to-ambiguity group (this is a relatively trivial finding, as it directly reflects the criterion used for our group split). More interestingly, the group split also resulted in a condition[*] group interaction effect for incoherence, as well as for combined ambiguity and incoherence: both the incoherent anaphors and the ambiguous/incoherent anaphors elicited a larger N400 effect in the no-LPC-to-ambiguity group than in the LPC-to-ambiguity group. This interaction was statistically significant for semantic incoherence, and marginally significant for ambiguity/incoherence. A related pattern of results was observed in the later time window: incoherent anaphors and also the ambiguous/incoherent anaphors elicited a smaller posterior LPC effect in the no-LPC-to-ambiguity group than in the LPC-to-ambiguity group.

At this point, it is important to mention that these group patterns cannot be an artefactual side-effect of dividing subjects on the basis of accidental EEG fluctuations ('noise'). Of course, it is well-known that with limited numbers of EEG trials going into

the average of a single subject, single-subject ERPs consist of an unknown but not necessarily favorable mixture of critical ERP effects and 'residual EEG background noise'. However, if our sorting of subjects on the basis of their ERPs to referentially ambiguous anaphors would primarily have been a sorting based on residual noise in this condition, then the resulting group-split should not have had such systematic consequences for the other conditions. The systematic covariance across conditions observed here shows that what we are looking at here must be something real.[3]

By clearing the view on both the Nref effect and the N400 effect in the no-LPC-to-ambiguity group, the present group split thus allowed for a more sensitive inspection of the interplay between semantic and referential analysis in the ambiguity/incoherence condition. In this group, ambiguous anaphors elicited a significant Nref effect (see Table 4). Incoherent and ambiguous/incoherent anaphors both elicited a significant N400 effect followed by a statistically significant LPC. Crucially, ambiguous/incoherent anaphors did not elicit any trace whatsoever of an Nref effect. Thus, the pattern of results across conditions for this selection of subjects was similar to that for all subjects pooled together.

In contrast, a completely different picture emerged for the LPC-to-ambiguity group. In that group, ambiguous anaphors only elicited an LPC effect (due to the group split), that was statistically significant in both time windows (see Table 5). Neither the semantically incoherent nor the ambiguous/incoherent anaphors elicited a reliable N400 effect, but both types of anaphors instead elicited a large LPC effect.

In a final post-hoc analysis, we directly tested for an additional Nref effect in the 500- to 1600-ms window following the ambiguous/incoherent anaphors compared to the incoherent anaphors. A pairwise comparison between the ERPs elicited by the incoherent anaphors and the ambiguous/incoherent anaphors revealed no significant differences across all subjects at either posterior electrodes (F(1,37) = 2.71, p = .11) or anterior electrodes (F(1,37) = .306, p = .583). Also no statistical differences occurred in the two groups separately: no-LPC-to-ambiguity group, posterior electrodes (F(1,19) = 1.68, p = .21), anterior electrodes (F(1,19) = .936, p = .345); LPC-to-ambiguity group, posterior electrodes (F(1,18) = 1.09, p = .311), anterior electrodes (F(1,18) = .01, p = .92). In short, even in the group where the Nref was most pronounced, the ERPs elicited by incoherent anaphors and by ambiguous/incoherent anaphors were statistically indistinguishable (i.e., no additive effect of referential ambiguity). Similarly, in the group where the LPC effects were most pronounced, there was no additive effect of referential ambiguity in the ERPs following the ambiguous/incoherent anaphors.

4. Discussion

In this event-related brain potential study, we examined the interplay between semantic and referential aspects of anaphoric noun phrase resolution in discourse comprehension. We used a full factorial design that crossed referential ambiguity with semantic incoherence. As expected, ambiguous anaphors elicited a sustained, frontal negative ERP effect that is typically associated with referential ambiguity (Nref effect; Van Berkum et al., 1999; Van Berkum et al., 2003), and incoherent anaphors elicited a standard N400 effect (Kutas & Hillyard, 1980). In addition, our study yielded two main novel findings. First, anaphors that were both ambiguous and incoherent elicited an ERP pattern resembling that of the

---

[3] We made a similar group split that involved the ERP responses to semantically incoherent anaphors, as some subjects showed only LPC effects and no N400 effects to these anaphors. Importantly, as these subjects generally belonged to the LPC-to-ambiguity group, this alternative group split resulted in a similar pattern of results as reported here. Out of parsimony, we do not report these additional results.
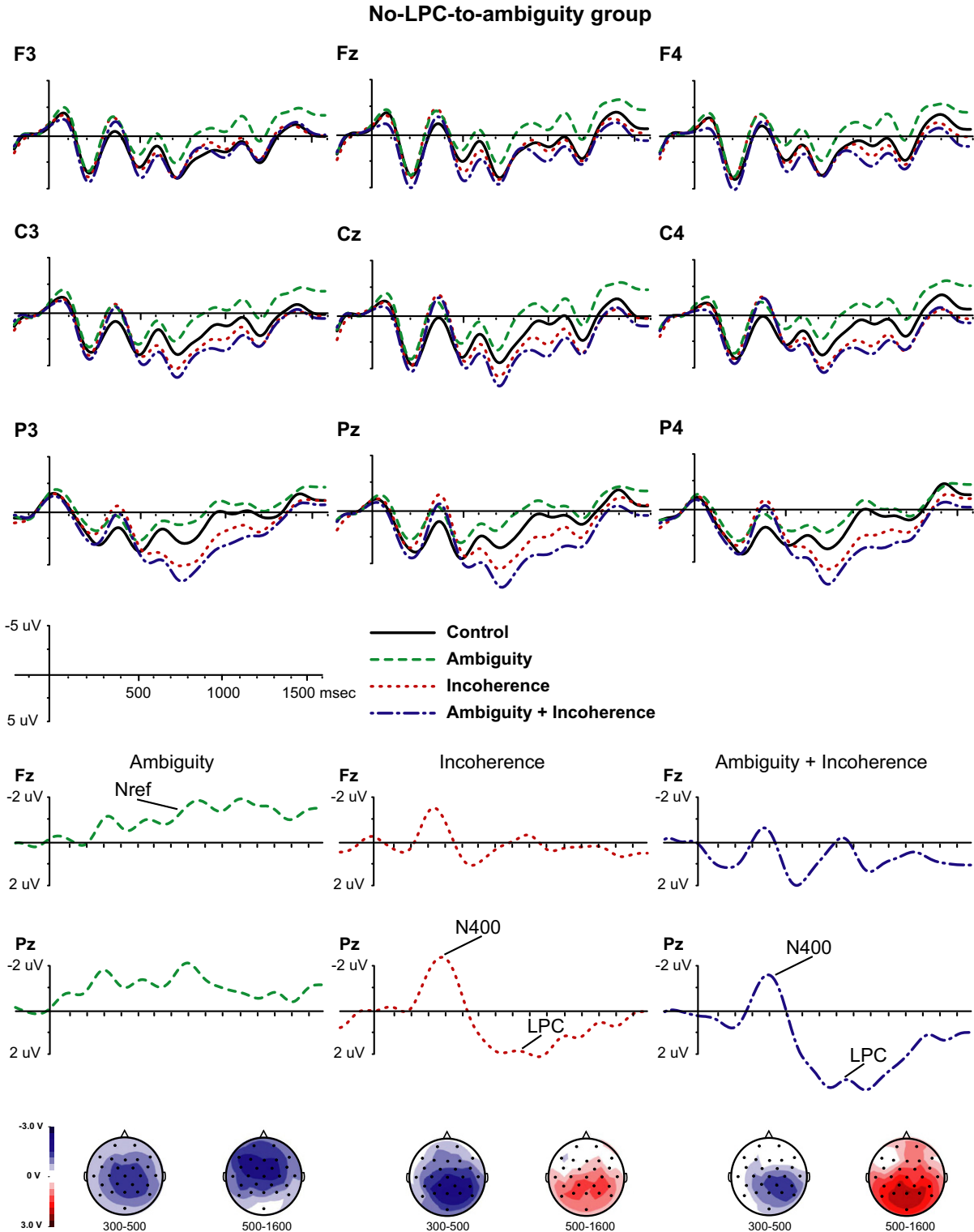
No-LPC-to-ambiguity group

Fig. 2. No-LPC-to-ambiguity group. (Upper graph) ERPs to referentially ambiguous anaphors, semantically incoherent anaphors, ambiguous/incoherent anaphors and the control anaphors at a subset of electrode positions. (Lower graph) Difference waves of ambiguity, incoherence and ambiguity/incoherence (compared to the control condition) at Fz and Pz, and corresponding scalp distributions for the two time windows that were used for statistical analyses.

incoherent anaphors. Although the N400 effect elicited by these anaphors was not statistically significant across all subjects, it was significant in the group of subjects that did show the expected effects of referential ambiguity and semantic incoherence separately. As we will argue below, these results can be taken to suggest that semantic incoherence can preclude readers from

taking into account two competing referential interpretations. Second, many of our subjects showed similar late positive ERP effects to referential ambiguity, semantic incoherence and to their combination. Importantly, the exclusion of these subjects (based on their LPC responses to referential ambiguity) had no effect on the general pattern of results. However, these LPC findings are interesting
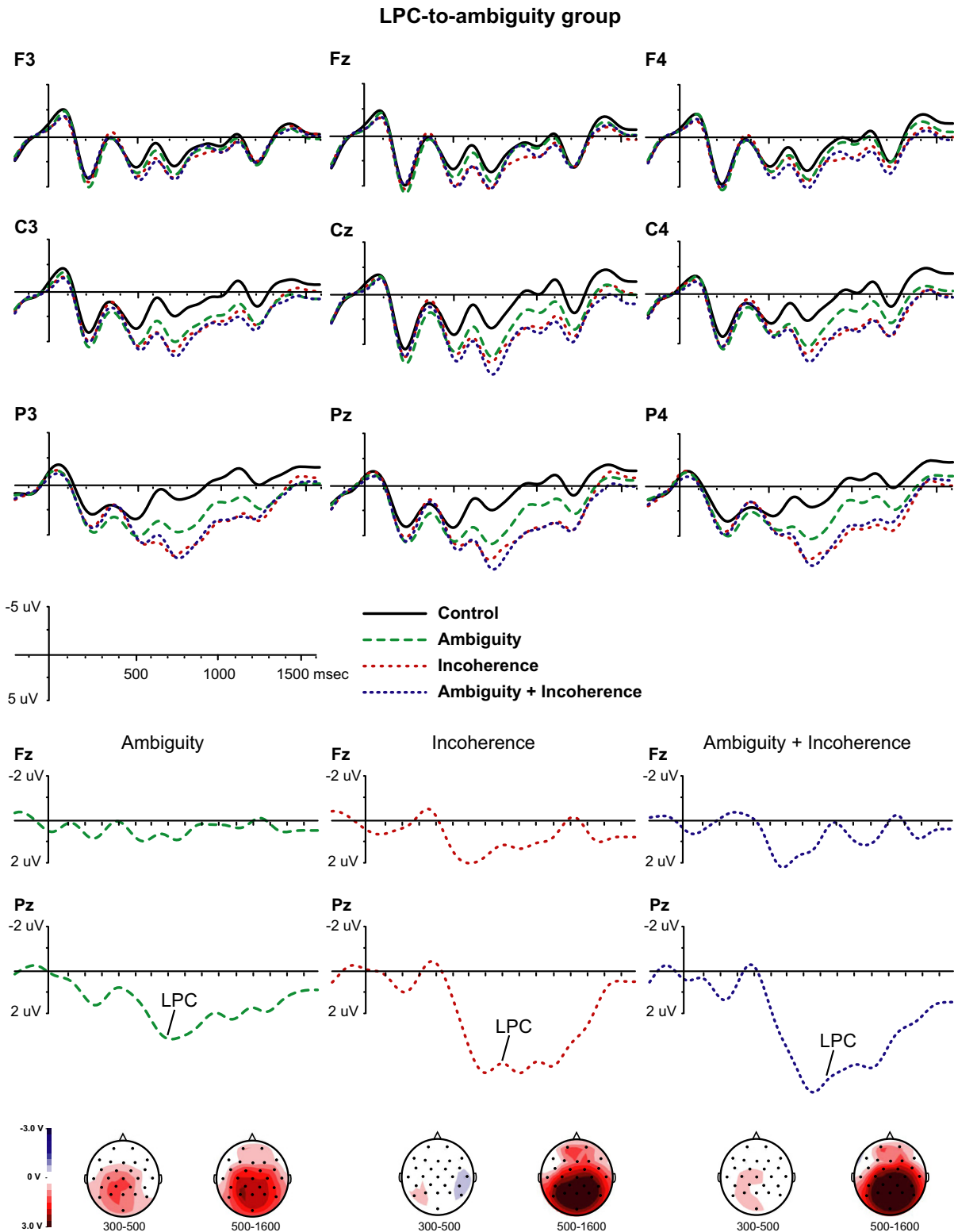
LPC-to-ambiguity group

Fig. 3. LPC-to-ambiguity group. (Upper graph) ERPs to referentially ambiguous anaphors, semantically incoherent anaphors, ambiguous/incoherent anaphors and the control anaphors at a subset of electrode positions. (Lower graph) Difference waves of ambiguity, incoherence and ambiguity/incoherence (compared to the control condition) at Fz and Pz, and corresponding scalp distributions for the two time windows that were used for statistical analyses.

in their own right and may reflect the extent to which participants, in addition to the 'normal' language comprehension task, processed for implausibility (i.e., paid special attention to whether the text conveys an event that is plausible in the real-world). In the below paragraphs, we will discuss these two findings and their implications.

4.1. Semantic incoherence precludes elaborate anaphoric inferencing

In our introduction, we distinguished several scenarios about how the language comprehension system would cope with anaphors that are simultaneously semantically incoherent and referentially ambiguous. We outlined the possibilities that semantic

Table 3
F, MSE, and *p* values for the condition * group interaction effects of ambiguity, incoherence, and their combination

	Electrodes	300–500			500–1600		
		F	MSE	p	F	MSE	p
Ambiguity	All	18.856	19.093	.000***	54.446	18.463	.000***
	Anterior	7.084	13.395	.011*	27.922	11.800	.000***
	Posterior	19.117	10.745	.000***	53.466	9.111	.000***
Incoherence	All	5.476	34.892	.025*	3.214	44.626	.081^
	Anterior	2.025	19.279	.163	1.072	26.406	.307
	Posterior	7.571	18.073	.009**	5.949	19.129	.020*
Combined	All	1.161	53.515	.288	1.543	70.994	.222
	Anterior	.033	25.918	.858	.153	23.367	.698
	Posterior	3.093	25.077	.087^	3.820	21.401	.058^

For all *F* tests, numerator *df* = 1, denominator *df* = 37, ^$p \leqslant$ .1, *$p \leqslant$ .05, **$p \leqslant$ .01, ***$p \leqslant$ .001.

Table 4
F, MSE, and *p* values for the main effects of ambiguity, incoherence, and their combination in the no-LPC-to-ambiguity group

	Electrodes	300–500			500–1600		
		F	MSE	p	F	MSE	p
Ambiguity	All	17.130	15.937	.001***	52.910	12.200	.000***
	Anterior	6.450	14.305	.020*	37.789	10.228	.000***
	Posterior	15.469	8.259	.001***	43.690	3.235	.000***
Incoherence	All	9.798	44.439	.006**	1.641	63.629	.216
	Anterior	4.161	21.557	.056^	.045	31.536	.834
	Posterior	15.140	20.657	.001***	5.783	25.096	.027*
Combined	All	2.512	26.740	.129	4.884	62.446	.040*
	Anterior	.026	13.344	.874	1.108	27.936	.306
	Posterior	6.081	15.696	.023*	11.222	27.497	.003**

For all *F* tests, numerator *df* = 1, denominator *df* = 19, ^$p \leqslant$ .1, *$p \leqslant$ .05, **$p \leqslant$ .01, ***$p \leqslant$ .001.

Table 5
F, MSE, and *p* values for the main effects of ambiguity, incoherence, and their combination in the LPC-to-ambiguity group

	Electrodes	300–500			500–1600		
		F	MSE	p	F	MSE	p
Ambiguity	All	4.816	22.424	.042*	15.182	25.074	.001***
	Anterior	1.446	12.433	.245	2.842	13.458	.109
	Posterior	6.055	13.370	.024*	24.157	15.314	.000***
Incoherence	All	.043	24.815	.838	28.977	24.568	.000***
	Anterior	.015	16.875	.904	3.517	20.990	.077^
	Posterior	.053	15.346	.820	55.312	12.829	.000***
Combined	All	.111	81.778	.742	28.883	28.691	.000***
	Anterior	.013	39.190	.911	3.508	18.543	.077^
	Posterior	.220	34.980	.644	59.131	14.965	.000***

For all *F* tests, numerator *df* = 1, denominator *df* = 18, ^$p \leqslant$ .1, *$p \leqslant$ .05, **$p \leqslant$ .01, ***$p \leqslant$ .001

incoherence would preclude referential ambiguity, that one or even both of these problems would be 'boosted' (e.g., Hagoort, 2003), or that referential processing would proceed independently of semantic processing. The fact that the ambiguous/incoherent anaphors elicit an ERP pattern closely resembling that of the incoherent anaphors lends support to the first scenario, and may suggest that semantic incoherence precludes the anaphoric inferences associated with referential ambiguity.

In earlier research, we have shown that referential ambiguity is also precluded by sentence- and discourse-level manipulations that render one of the referential candidates to be an implausible antecedent (Nieuwland & Van Berkum, 2006; Nieuwland et al., 2007a; see also Tanenhaus & Trueswell, 1995). For example, *nephew* did not elicit a referentially induced ERP effect in stories where two nephews were originally introduced but one of which was unavailable as a referential candidate (for example, one nephew had died). In the current study, however, both potential antecedents were equally (un-)likely, and we propose a different mechanism for why readers did not take into account both candidates.

This hypothesis is based on the notion that human language comprehension is limited by processing resources availability. Note that semantically incoherent anaphors constitute direct violations of semantic world knowledge, whereas referentially ambiguous anaphors are not unexpected or anomalous per se (rather, they are incomplete). Semantic incoherence therefore constitutes a stronger, hence more salient disruption of a readers' standard of coherence than referential ambiguity (e.g., see Nieuwland & Van Berkum, 2006, who showed that subjects may greatly differ in whether they perceive 'formally' ambiguous pronouns to genuinely ambiguous, even when they elicit identical responses to semantic anomalies). The fact that the ambiguous/incoherent anaphors elicited an ERP pattern resembling that to the incoherent anaphors suggest that readers set out to work on the most salient and important coherence break. Because our language comprehension system has only limited processing capacity, this focus on re-establishing semantic coherence diverted readers' attention from the 'smaller concern' imposed by referential ambiguity.[4] As such, the current results provide convergent evidence that engaging in referential ambiguity resolution (as reflected by the referentially induced ERP effect) involves elaborate anaphoric inferencing (Nieuwland et al., 2007b; Nieuwland & Van Berkum, 2006; see also Greene et al., 1992).

The present ERP results can be taken to complement recent behavioral results on anaphoric inferencing. Levine and colleagues showed that if a readers' standard of coherence (Van den Broek et al., 1995) is already met without fully reinstating an antecedent, it is inefficient for readers to engage in elaborate anaphor resolution, and anaphoric inferences are optional rather than necessary (e.g., Levine et al., 2000; Klin et al. 2004; Klin et al. 2006). For example, when a salient distractor (e.g., 'cake') intervened between an anaphoric noun phrase (e.g., 'the dessert') and its antecedent (e.g., 'tart'), readers failed to reinstate the antecedent (as measured by recognition probe response times as well as reading times; Levine et al., 2000). Although drawing the anaphoric inference could have enhanced comprehension of the discourse, readers failed to identify a specific antecedent because the anaphor could plausibly refer to a non-specific antecedent (i.e., there was no coherence break). Follow-up results showed that saliency-enhancing manipulations (e.g., bringing the anaphor into syntactic focus, or adding pronominal adjectival modifiers) caused readers to engage in anaphoric inferencing (Klin et al., 2004).

The present results suggest that semantic violations are also able to modulate the allocation of attentional resources to anaphoric inferencing. In this respect, our results suggest that readers can fail to engage in elaborate anaphoric inferencing not only because making such inferences can be superfluous in an already fully coherent discourse (Levine et al., 2000), but also because making these inferences can be of secondary importance in an already *incoherent* discourse. As mentioned in the introduction, this is not a situation that readers will encounter on a regular basis in everyday

---

[4] Note that such an idea also receives indirect support from our functional neuroimaging results. We have reported a double dissociation between semantic incoherence and referential ambiguity in their respective MRI BOLD responses (Nieuwland et al., 2007b), which could be taken to suggest that (some aspects of) semantic and referential analysis are mutually dependent in terms of information processing streams and/or neural processing resources.

life. At the same time, however, it does clearly indicate that one aspect of conceptual interpretation, semantic analysis, can proceed when another aspect of conceptual processing, referential analysis, cannot deliver a unique referent over which semantic analysis is ultimately defined.

Note that we do not claim that semantic incoherence necessarily precludes other aspects of referential processing. As referential processing can be intricately intertwined with semantic or syntactic processing, referential problems (e.g., overly specific anaphors, referential failure) will greatly vary in their perceived saliency, and therefore in the amount of processing resources that may be devoted to overcoming a coherence break. Similarly, it remains to be seen whether our results will replicate using semantic coherence breaks that do not violate semantic world knowledge but only involve slightly unexpected scenarios. Although initial steps have been made (e.g., Levine et al., 2000; Klin et al., 2004, 2006), future research should elucidate the different circumstances that cause readers to fail to resolve the 'full' referential implications of anaphoric expressions. In any case, the current results suggest that re-establishing semantic coherence has functional precedence over at least one aspect of referential processing, namely elaborate anaphoric inferencing.

4.2. Common LPC effects to linguistically different problems

Our split group analysis showed that whereas one group of subjects showed the expected ERP effects of referential ambiguity and semantic incoherence, another group predominantly showed late positive ERP effects to referential ambiguity, semantic incoherence as well as to their combination. Although the latter LPC-to-ambiguity group elicited somewhat smaller LPC effects to ambiguity compared to incoherence and ambiguity/incoherent, these effects had a very similar scalp distribution and morphology across the three conditions.

It is important to note that, whereas only the no-LPC-to-ambiguity group showed straightforward and significant N400 effects to incoherence and ambiguity/incoherence, the LPC-to-ambiguity group did reveal traces of N400 effects as well. These N400 traces are best discernible in the difference waveforms as negative deflections on top of the larger LPC effects, having a similar time course and morphology as the N400 effects in the no-LPC-to-ambiguity group. This suggests the possibility that the LPC-group-to-ambiguity group did generate the expected N400 effects, but that these effects were effectively 'masked' by the large LPC effects in this group. However, because the Nref effect that we usually observe to referential ambiguity is a small, morphologically less 'distinct' (e.g., less peaked) effect that overlaps with the present LPC effects in time and distribution, it is more difficult to determine whether this group, next to its large LPC, also displays an Nref effect. However, we cannot rule out this possibility, and must accept that the data are inconclusive on this point.

Also relevant to this issue is the observation is that the no-LPC-to-ambiguity group elicited LPC effects to incoherence and ambiguity/incoherence, in addition to the N400 effects.[5] These LPC effects were comparable to the LPC effects in the LPC-to-ambiguity group in terms of scalp distribution and morphology. Furthermore, in the LPC-to-ambiguity group, the stronger violations imposed by the incoherent and ambiguous/incoherent anaphors elicited larger LPC effects than the less disruptive ambiguous anaphors. In all, the most parsimonious account for our complete set

of results seems to be that the two groups responded *qualitatively* similar in terms of the pattern of N400, LPC, and — perhaps — sustained negative shift effects,[6] but that the groups did differ in the *size* of their LPC responses. Note that earlier language comprehension studies have already uncovered large individual differences in anomaly-induced late positivities (e.g., Osterhout, 1997; Osterhout & Mobley, 1995).

Clearly, the important question remains what processes are indexed by the current LPC response, and what causes its variability across individuals. Late positive ERP effects in language comprehension paradigms have traditionally been associated with syntactic analysis or re-analysis (the P600; Osterhout & Holcomb, 1992), but this strictly syntactic account seems hard to reconcile with the current results (see also Muente, Heinze, Matzke, Wieringa, & Johannes, 1998; Osterhout, 1997). The literature on late positive ERP effects in language comprehension paradigms suggests at least three other options.

First, the current late positive ERP effects may be related to the P300 ERP response, a domain-general response to task-relevant events (Coulson, King, & Kutas, 1998; Pritchard, 1981; but see Osterhout & Hagoort, 1999; Osterhout, McKinnon, Bersick, & Corey, 1996). Although we did not give participants an explicit task, it is possible that they, in addition to the 'normal' task of language comprehension, perceived the anaphors to be relevant to an implicit task (e.g., judging the acceptability of anaphoric expressions).

The other options involve two recent functional accounts of the P600 that were formulated to account for unexpected P600 effects following semantic violations (e.g., Kim & Osterhout, 2005; Kolk & Chwilla, 2007; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). According to Kolk and colleagues, P600 effects reflect a general kind of reprocessing aimed at editing out misperceptions or processing errors of different kinds (the monitoring hypothesis; e.g., Kolk & Chwilla, 2007). An alternative, but related account (Kuperberg, 2007) holds that P600 effects are elicited by the conflict between semantic memory-based and combinatorial processing streams (syntactic, semantic–thematic), and reflect continued analysis within the particular combinatorial processing stream that detected the conflict (instead of generalized reprocessing across levels of analysis).

The current data do not allow us to distinguish between these three different accounts, and what the current LPC effects really index is not limited to the abovementioned options. However, the three accounts do commonly suggest that the LPC variability across individuals potentially lies in subject task strategies. In the P300 account, individual variability in the obtained LPC effects could reflect the extent to which participants adopted an implicit task. In a monitoring account (e.g., Kolk & Chwilla, 2007), the present LPC variability could reflect the extent to which participants were differentially monitoring for processing errors, whereas in a 'conflict' account, it could reflect the variable 'tuning' of semantic memory-based and combinatorial processing mechanisms (Kuperberg, 2007). Both the latter options assume an important role of task strategy as modulating executive control over language comprehension, be it over the monitoring process or over the balance between processing streams.

The notion that individual processing strategies may underlie the variability in current LPC effects is consistent with the task-dependency of late positive ERP effects following semantic violations. Semantic violations particularly elicit late positive ERP effects when subjects are required to judge sentences for implausi-

---

[5] In fact, it is possible that the LPC effects also counteracted the N400 effects in this group to some extent, which could explain why we obtained relatively small N400 effects (~2 μV amplitude difference). Note that some N400 studies in the experimental literature have reported subsequent LPC effects. A dedicated investigation into these later effects is yet to be done.

[6] It would be interesting, however, if the two groups genuinely differed in their Nref response to referential ambiguity, which could reflect qualitative differences in individual processing strategies (e.g., upon noticing referential ambiguity, some subjects engaged in anaphoric inferences to 'search' the correct antecedent whereas others did not).

bility or unacceptability (e.g., Kuperberg, 2007; see also Osterhout & Mobley, 1995, for effects of task strategy on 'syntactic' P600 effects), whereas these effects can disappear when subjects are instructed to ignore implausibility (Vissers, Chwilla, & Kolk, 2007). In the light of these findings, one can speculate that the present group differences reflect the extent to which the groups, on top of the other comprehension processes, were processing for implausibility or unacceptability. If this hypothesis is correct, the currently obtained ERP patterns should be sensitive to an instruction-manipulation (e.g., Vissers et al., 2007), or to the addition of a plausibility judgment task (e.g., Kuperberg, 2007).

In sum, we have no conclusive evidence to the exact functional significance of the common LPC effects across groups and conditions. However, the experimental literature suggests that they may reflect the extent to which participants, in addition to the 'normal' language comprehension task, processed for implausibility. Importantly, and irrespective of what the common LPC effects entail, we were able to address the main focus of the present study, as our results provide new insights into the interplay between semantic and referential aspects of anaphor resolution. In the group of subjects where the ERP effects associated with referential ambiguity and semantic incoherence were not obscured by LPC effects, the combination of ambiguity and incoherence elicited an ERP pattern resembling that to incoherence separately.

5. Conclusions

We examined the interplay between semantic and referential aspects of anaphoric noun phrase resolution, using a full factorial design that crossed referential ambiguity with semantic incoherence. Anaphors in the critical condition, which combined referential ambiguity with semantic incoherence, elicited ERPs resembling that to the separate semantically incoherent anaphors. These results suggest that re-establishing semantic incoherence precludes readers from engaging in elaborate anaphoric inferencing. Additionally, a subset of participants showed unexpected late positive effects to all three types of problematic anaphors, suggesting that people differ in the extent to which they, on top of the other comprehension processes, process for implausibility. This finding may have implications for recent debates on the functional significance of the P600 effect.

Acknowledgments

Appendix A

Additional example stories (approximate English translations from Dutch, and original Dutch stories)

A.1. Example story 1

A.1.1. Semantically coherent + referentially unique control
Police officer Laura had arrested a pushy thief and a sneaky voyeur in the Vondelpark. Laura brought both men to the police headquarters. She arranged a cell for the voyeur so the park would remain safe.

Agente Laura had in het Vondelpark een opdringerige dief en een stiekeme voyeur gearresteerd. Laura bracht beide mannen naar het hoofdbureau. Ze regelde een cel voor de voyeur zodat het park veilig bleef.

A.1.2. Referentially ambiguous
Police officer Laura had arrested two voyeurs in the Vondelpark, a pushy one and another one that was sneaky. Laura brought both men to the police headquarters. She arranged a cell for the voyeur so the park would remain safe.

Agente Laura had in het Vondelpark twee voyeurs gearresteerd, een stiekeme en de ander een opdringerige. Laura bracht beide mannen naar het hoofdbureau. Ze regelde een cel voor de voyeur zodat het park veilig bleef.

A.1.3. Semantically incoherent
Police officer Laura had arrested a pushy thief and a sneaky voyeur in the Vondelpark. Laura brought both men to the police headquarters. She locked them up in the voyeur so the park would remain safe.

Agente Laura had in het Vondelpark een opdringerige dief en een stiekeme voyeur gearresteerd. Laura bracht beide mannen naar het hoofdbureau. Ze sloot ze op in de voyeur zodat het park veilig bleef.

A.1.4. Semantically incoherent + referentially ambiguous
Police officer Laura had arrested two voyeurs in the Vondelpark, a pushy one and another one that was sneaky. Laura brought both men to the police headquarters. She locked them up in the voyeur so the park would remain safe.

Agente Laura had in het Vondelpark twee voyeurs gearresteerd, een stiekeme en de ander een opdringerige. Laura bracht beide mannen naar het hoofdbureau. Ze sloot ze op in de voyeur zodat het park veilig bleef.

A.2. Example story 2

A.2.1. Semantically coherent + referentially unique control
Bob wanted to discover his spiritual self in India, and when he was there met a near-sighted guru and a fairly deaf shaman. Bob found both men to be extremely inspiring. As advised by the guru he read books about meditation.

Bob wilde in India zijn spiritualiteit ontdekken en ontmoette daar een bijziende goeroe en een slechthorende sjamaan. Bob vond beide mannen uitermate inspirerend. Op advies van de goeroe las hij boeken over meditatie.

A.2.2. Referentially ambiguous
Bob wanted to discover his spiritual self in India, and when he was there met two gurus, a near-sighted one and also a fairly deaf one. Bob found both men to be extremely inspiring. As advised by the guru he read books about meditation.

Bob wilde in India zijn spiritualiteit ontdekken en ontmoette daar twee goeroes, de een was bijziend en de andere was slechthorend. Bob vond beide mannen uitermate inspirerend. Op advies van de goeroe las hij boeken over meditatie.

A.2.3. Semantically incoherent
Bob wanted to discover his spiritual self in India, and when he was there met a near-sighted guru and a fairly deaf shaman. Bob found both men to be extremely inspiring. While travelling through the guru he read books about meditation.

Bob wilde in India zijn spiritualiteit ontdekken en ontmoette daar een bijziende goeroe en een slechthorende sjamaan. Bob vond beide mannen uitermate inspirerend. Op de doorreis in de goeroe las hij boeken over meditatie.

A.2.4. Semantically incoherent + referentially ambiguous
Bob wanted to discover his spiritual self in India, and when he was there met two gurus, a near-sighted one and also a fairly deaf one. Bob found both men to be extremely inspiring.

While travelling through the guru he read books about meditation.

Bob wilde in India zijn spiritualiteit ontdekken en ontmoette daar twee goeroes, de een was bijziend en de andere was slechthorend. Bob vond beide mannen uitermate inspirerend. Op de doorreis in de goeroe las hij boeken over meditatie.

A.3. Example story 3

A.3.1. Semantically coherent + referentially unique control

A funny brunette and a mysterious blonde made a pass at Eric in the disco-club. Eric thought both women were cute and attractive. He started dancing with the brunette to make a good impression.

Erik werd in de disco versierd door een grappige brunette en een mysterieuze blondine. Erik vond beide vrouwen leuk en aantrekkelijk. Hij begon te dansen met de brunette om indruk te maken.

A.3.2. Referentially ambiguous

Two brunettes made a pass at Eric in the disco-club, one was funny and the other was mysterious. Eric thought both women were cute and attractive. He started dancing with the brunette to make a good impression.

Erik werd in de disco versierd door twee brunettes, de een was grappig en de ander was mysterieus. Erik vond beide vrouwen leuk en aantrekkelijk. Hij begon te dansen met de brunette om indruk te maken.

A.3.3. Semantically incoherent

A funny brunette and a mysterious blonde made a pass at Eric in the disco-club. Eric thought both women were cute and attractive. He started dancing on the brunette to make a good impression.

Erik werd in de disco versierd door een grappige brunette en een mysterieuze blondine. Erik vond beide vrouwen leuk en aantrekkelijk. Hij begon te dansen op de brunette om indruk te maken.

A.3.4. Semantically incoherent + referentially ambiguous

Two brunettes made a pass at Eric in the disco-club, one was funny and the other was mysterious. Eric thought both women were cute and attractive. He started dancing on the brunette to make a good impression.

Erik werd in de disco versierd door twee brunettes, de een was grappig en de ander was mysterieus. Erik vond beide vrouwen leuk en aantrekkelijk. Hij begon te dansen op de brunette om indruk te maken.

References

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

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Auer, J. C. P. (1984). Referential problems in conversation. *Journal of Pragmatics, 8*, 627–648.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Reviews of Neurosciences, 25*, 151–188.

Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 881–895). Cambridge, MA: MIT Press.

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

Chierchia, G., & McConnell-Ginet, S. (1990). *Meaning and grammar: An introduction to semantics*. Cambridge, MA: MIT Press.

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

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 121–132.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors. *Language and Cognitive Processes, 22*(6), 793–827.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Philadelphia, PA: Taylor and Francis.

Garnham, A., Oakhill, J., & Cain, K. (1997). The interpretation of anaphoric noun phrases: Time course, and effects of overspecificity. *The Quarterly Journal of Experimental Psychology, 50A*, 149–162.

Garrod, S. C., & Sanford, A. J. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, 16*, 77–90.

Garrod, S., & Sanford, A. J. (1981). Bridging inferences in the extended domain of reference. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 331–346). Hillsdale, NJ: Erlbaum.

Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). San Diego, CA, USA: Academic Press.

Garrod, S. C., & Sanford, A. J. (1999). Incrementality in discourse understanding. In Herre van Oostendorp & R. Goldman Susan (Eds.), *The construction of mental representations during reading* (pp. 3–27). Mahwah, NJ, USA: Lawrence Erlbaum.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32*, 99–156.

Gibbs, R. W. (1984). Literal meaning and psychological theory. *Cognitive Science, 8*, 275–304.

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

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 266–283.

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

Jackendoff, R. (2002). *Foundations of language*. New York, NY: Oxford University Press.

Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205–225.

Klin, C. M., Guzman, A. E., Weingartner, K. M., & Ralano, A. S. (2006). When anaphor resolution fails: Partial encoding of anaphoric inferences. *Journal of Memory and Language, 54*, 131–143.

Klin, C. M., Weingartner, K. M., Guzmán, A. E., & Levine, W. H. (2004). Readers' sensitivity to linguistic cues in narratives: How salience influences anaphor resolution. *Memory & Cognition, 32*, 511–522.

Kolk, H. H. J., & Chwilla, D. J. (2007). Late positivities in unusual situations. *Brain and Language, 100*(3), 257–261.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1–36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117–129.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

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

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA, USA: Academic Press, Inc..

Ledoux, K., Gordon, B., Roll, M., & Swaab, T. (2007). The influence of implicit causality on the establishment of coreference: An event-related potential study of reading. *Annual meeting of the cognitive neuroscience society (CNS-2007)*, May 5–9, New York.

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading. XVI. The visual span in normal and low vision. *Vision Research, 37*(14), 1999–2010.

Levinson, S. C. (2000). *Presumptive meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: The MIT Press.

Levine, W. H., Guzmán, A. E., & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language, 43*, 594–617.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America, 94*(20), 10979–10984.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19*, 668–682.

Muente, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217–226.

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

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007a). Who are you talking about? Tracking discourse-level referential processing with ERPs. *Journal of Cognitive Neuroscience, 19*, 228–236.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007b). On sense and reference: Examining the functional neuroanatomy of referential processing. *Neuroimage, 37*(3), 993–1004.

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

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

O'Brien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes, 23*, 1–24.

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

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*, 1–14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8*, 507–526.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 283–317.

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research, 11*(1153), 166–177.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin, 89*(3), 506–540.

Sanford, A. J., & Garrod, S. C. (1989). What, when, and how? Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes, 4*(3-4), SI235–SI262.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*(9), 382–386.

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

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.). *Speech, language, and communication* (pp. 217–262). San Diego, CA, USA: Academic Press.

Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions.* Cambridge, MA: MIT Press.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 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*(2), 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 Research, 1146*, 158–171.

Van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–374). Hillsdale, NJ: Erlbaum.

Vissers, C. Th. W. M., Chwilla, D. J., & Kolk, H. H. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology, 75*(1), 8–18.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences, 9*(9), 445–453.

Walker, F., & Yekovich, C. (1987). Activation and use of script-based antecedents in anaphoric reference. *Journal of Memory and Language, 26*, 673–691.

Wilson, D., & Sperber, D. (2004). Relevance theory. In G. Ward & L. Horn (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford: Blackwell.