Research Report

Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension

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

[a]Department of Psychology (PN), University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands
[b]F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

Abstract

In general, language comprehension is surprisingly reliable. Listeners very rapidly extract meaning from the unfolding speech signal, on a word-by-word basis, and usually successfully. Research on 'semantic illusions' however suggests that under certain conditions, people fail to notice that the linguistic input simply doesn't make sense. In the current event-related brain potentials (ERP) study, we examined whether listeners would, under such conditions, spontaneously detect an anomaly in which a human character central to the story at hand (e.g., "a tourist") was suddenly replaced by an inanimate object (e.g., "a suitcase"). Because this replacement introduced a very powerful coherence break, we expected listeners to immediately notice the anomaly and generate the standard ERP effect associated with incoherent language, the N400 effect. However, instead of the standard N400 effect, anomalous words elicited a positive ERP effect from about 500–600 ms onwards. The absence of an N400 effect suggests that subjects did not immediately notice the anomaly, and that for a few hundred milliseconds the comprehension system has converged on an apparently coherent but factually incorrect interpretation. The presence of the later ERP effect indicates that subjects were processing for comprehension and did ultimately detect the anomaly. Therefore, we take the absence of a regular N400 effect as the online manifestation of a temporary semantic illusion. Our results also show that even attentive listeners sometimes fail to notice a radical change in the nature of a story character, and therefore suggest a case of short-lived 'semantic change deafness' in language comprehension.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Under normal circumstances, the human language comprehension system works amazingly fast, and amazingly well. One of the key features of the system that allows for its usually excellent performance is incrementality. Psycholinguistic experiments have shown that listeners and readers immediately relate the meaning and grammar of each incoming word to the context, whether this context consists of an isolated sentence (e.g., [28,51]), the global discourse (e.g., [17,29,43–45]), or a non-linguistic visual scene (e.g., [3,37]).

But immediate analysis does not necessarily mean that linguistic information is always interpreted to the fullest degree possible. For instance, there are quite a few reports of incomplete semantic analysis or underspecification (e.g., [4,10]; for review, see [12,36]). Particularly striking are semantic illusions, which indicate that sometimes the full meaning of a word is not incorporated into the interpretation of a sentence, with people happily accepting an incorrect interpretation based on semantic heuristics instead. A well-known example is the "Moses illusion" [10], in which participants routinely fail to notice the anomaly in the

* Corresponding author. Fax: +31 20 6391656.
  E-mail address: m.s.nieuwland@uva.nl (M.S. Nieuwland).

question '*How many animals of each sort did Moses put on the ark?*', despite knowing it was Noah instead of Moses. Similarly, after reading a story about a plane crash and confronted with the sentence '*The authorities had to decide where to bury the survivors*', participants regularly do not perceive the coherence break [4].

The available evidence suggests that the probability of detecting a semantic anomaly decreases if the impostor word (e.g., "Moses") is semantically related to the correct word ("Noah"; [10]), as well as if the impostor word is strongly associated with the global scenario suggested by the discourse (e.g., "survivors" in a plane crash scenario; [4]). Furthermore, we know that impostor words are more easily detected when they are brought into focus (e.g., as in 'It was Moses who put two of each kind of animal on the ark. True or False?', [5]), indicating that semantic illusions are partially dependent on the misdirection of focus. To account for these phenomena, Sanford and Sturt [36] have proposed that if a word fits the global situation very well or when it is out of focus, it will receive incomplete semantic analysis, to such an extent that the actual input can be misconstrued and a semantic anomaly can pass undetected.

On the one hand, the existence of semantic illusions is perhaps not that amazing. After all, the heavy cognitive demands imposed by processing speech or written text at a rate of several words per second makes it very unlikely that people always exploit every single bit of relevant information to the fullest degree. We know that in other domains of cognition, our brain often employs plausibility strategies, so that it can get where it needs to be quickly, while maintaining sufficient accuracy [34]. In decision making, for instance, our brain makes heavy use of heuristics that usually work well but inevitably go wrong from time to time (e.g., [20]). As evidenced by many classic visual illusions, even our highly esteemed visual system sometimes gets it wrong. Recent demonstrations of visual 'change blindness' (e.g., [38]), in which observers fail to notice a change to central objects in a scene even when looking for it, forcefully illustrate the fact that our input systems are not designed to consistently deliver full and complete representations of the input they encounter. Language comprehension may be no different. In fact, semantic illusions in language comprehension have recently been related to the use of a plausibility heuristic, which biases the system towards semantic analyses that are most consistent with world knowledge (e.g., [12,36]).

From another perspective, however, the existence of these illusions is actually very surprising. After all, story characters are central to situation models, and readers appear to be intensively engaged in keeping track of such 'protagonists' during comprehension [54]. If so, then why do people allow Moses to replace Noah without a blink, and allow survivors to replace protagonists who actually died? If protagonists are so central to the situation model, wouldn't one expect that listeners always immediately notice a protagonist being replaced by an entirely different entity,

even if the anomalous impostor word is out of prosodic focus and it fits a scenario very well?

This raises the issue of just how far we can stretch the limits of the semantic illusion phenomenon. For instance, would listeners, given the above mentioned conditions, fail to notice cases in which a salient human story character (e.g., a tourist) engaged in conversation with another human being (say, a check-in desk clerk) is suddenly replaced by a non-living entity (e.g., a suitcase) that happily continues the conversation? Animacy is a core semantic feature that lies at the heart of our understanding of the world around us. In fact, the animate–inanimate distinction is often considered to be an innate organizing principle of cognition (e.g., [14,34]), with different neural mechanisms subserving each of the two categories [7]. Even if an inanimate impostor word such as "suitcase" would be prosodically unfocused, scenario-relevant, and semantically related to the animate character being replaced, it does lack a core semantic feature required in this context, namely, being alive. Would even such a flagrant violation remain undetected?

We explored this issue in the present event-related brain potential (ERP) study by investigating whether participants would detect a discourse anomaly of the type just outlined under conditions that have been reported to induce semantic illusions. Participants listened to short narratives portraying a man and a woman engaged in a conversation about some inanimate object (see Table 1). The inanimate object and the man were semantically related to each other (e.g., 'diver' and 'harpoon', 'coachman' and 'whip'). A replacement occurred in the fifth and following sentences: Either the woman would continue her conversation with the man as usual (*coherent continuation*) or she would suddenly address the inanimate object instead and continue her conversation with this entity (*anomalous continuation*). Equivalent to the coherent continuation, the anomalous continuation was prosodically de-accented, and as such did

Table 1
Example story (approximate translation from Dutch)

Introduction

A tourist wanted to bring his huge suitcase onto the airplane. However, because the suitcase was so heavy, the woman behind the check-in counter decided to charge the tourist extra. In response, the tourist opened his suitcase and threw some stuff out. So now, the suitcase of the resourceful tourist weighed less than the maximum twenty kilos.

Coherent continuation	*Anomalous continuation*
Next, the woman told the *tourist* that she thought he looked really trendy. The tourist grabbed the woman's hand and eagerly asked her for a date. But the woman reprimanded the tourist for being pushy and told him to just get on the plane right away.	Next, the woman told the *suitcase* that she thought he looked really trendy. The suitcase grabbed the woman's hand and eagerly asked her for a date. But the woman reprimanded the suitcase for being pushy and told him to just get on the plane right away.

not focus attention on the impostor word. In spite of its prosodic 'low profile', though, the impostor word did introduce a flagrant anomaly, since it resulted not only in a strong discourse coherence break but also in a mismatch between animacy requirements imposed by a specific transitive verb and the inanimate argument actually supplied to this verb.

To assess whether our subjects would detect the anomaly, we relied on the N400 [24], a negative deflection in the ERP waveform that is elicited by every content word of an unfolding sentence, and peaks at approximately 400 ms after word onset. The amplitude of the N400 is known to be highly sensitive to the ease with which the meaning of the word at hand can be integrated into the prior sentence or discourse context. Semantically anomalous words, for instance, elicit a larger N400 than semantically coherent words (e.g., [24,44]; see [6,23,26] for reviews). The N400 effect is also elicited by words that are fully coherent but somewhat less expected (e.g., [16,25,39,46]). It is because of the latter findings that the N400 effect is not viewed as an anomaly marker, but is instead generally taken to reflect semantic integration processes that *occur routinely as words and sentences unfold*, as part of everyday language comprehension.

With spoken language comprehension, N400 effects typically begin to emerge at 150–250 ms after acoustic onset of the critical word (e.g., [11,44,51]). This rapid onset is taken to reflect the immediacy with which every unfolding spoken word is routinely related to its semantic context. As revealed by recent analyses involving spoken-language N400 effects (e.g., [47,48,51]), listeners in fact routinely relate an unfolding word to its context *before the word has become acoustically unique*, i.e., before enough of the word has been heard for the listener to know exactly which word it is going to be. These N400 findings converge with findings from eye-tracking research, which have revealed that listeners also extremely rapidly relate the meaning of an unfolding word to a visual scene (e.g., [3,37]).

Numerous studies have demonstrated language-elicited N400 effects under a variety of conditions and experimental manipulations. Two findings are particularly relevant to our present experimental logic. One is that the N400 effect, originally and most commonly established in isolated sentences, is equally sensitive to the ease with which a word can be related to the wider discourse. This has been demonstrated with discourse-dependent anomalies (e.g., [30,44,45]) as well as with words whose discourse-semantic fit differs in more subtle ways (e.g., [39,46]). Importantly, the timing of a discourse-dependent N400 effect is the same as that of its classic sentence-dependent counterpart, typically emerging in the ERP record within some 150–250 ms after acoustic or visual onset of the critical word, and invariably culminating in a large amplitude difference in the 300–500 ms interval.

The second N400 effect that is particularly relevant to our logic is one that we recently obtained in response to animacy violations involving the same set of critical words. In a spoken-language experiment related to the present study [30], participants listened to stories in which the *first* sentence occasionally contained a verb–argument animacy mismatch (e.g., "The woman spoke to the suitcase"). In such sentential contexts, inanimate nouns elicited a standard and large N400 effect relative to an animate control noun (e.g., "tourist"), emerging at about 250 ms after acoustic word onset (see also [52]). Thus, when listeners encounter sentences like "The woman spoke to the suitcase" without any further context, they immediately find out about the anomaly, as part of their routine process of semantic integration in language comprehension.

The critical issue addressed in the present study was whether or not participants would notice the very same anomaly, embedded in similar sentences, under conditions that have been reported to lead to semantic illusions. Based on the findings just discussed, and given the radical nature of the discourse-level coherence break, we expected our listeners to immediately notice the anomaly and as such generate a standard N400 effect in the 300–500 ms latency range. On the other hand, if under the 'semantic illusion' conditions investigated here listeners in fact fail to immediately detect the coherence break, no such N400 effect should be obtained.

2. Method

2.1. Participants

Thirty-five right-handed college students (15 men, mean age 21.4) participated in this study for course credit. All participants were native speakers of Dutch and without any neurological impairment.

2.2. Stimuli

Participants listened to 60 naturally spoken Dutch stories. To allow us to record the EEG without an attention-recruiting and potentially biasing secondary task (see [42]), we designed these stories to be as engaging as possible under the constraints of the present design. Each of the stories consisted of 7 sentences and contained three different main entities. Those entities were always one woman, one scenario-relevant man, and one scenario-relevant inanimate object. The nouns denoting the men and inanimate objects were semantically related to each other, and closely matched for word duration (mean 515 and 516 ms, respectively) and word form frequency (mean 1.78 and 1.75 on a million; Spoken Dutch Corpus). To provide a benchmark for our results, we used critical words that in an earlier study [30] had elicited a reliable N400 effect in story-initial sentences that contained similar animacy violations (e.g., "the girl told the harpoon").

The first 4 sentences were identical across conditions (see Table 1): the man and woman were engaged in conversation, and the inanimate object constituted the theme of the story. Each of these sentences contained a noun phrase anaphor to both the man and the inanimate object, matching our critical words for earlier mention. In the fifth sentence, the two conditions diverged. Either the woman would continue her conversation with the man (*coherent continuation*) or she would suddenly start talking to the inanimate object instead (*anomalous continuation*). The last 3 sentences were identical across conditions except for the critical words. For each story a coherent and anomalous version were recorded with a normal speaking rate and intonation, by the same female native speaker. The speaker was requested to minimize prosodic differences across the two versions of a story. In particular, the critical impostor word (e.g., 'suitcase') was always de-accented in a way that would be appropriate for the 'given' entity (e.g., 'tourist'). If necessary, recordings were redone to achieve this.

2.3. EEG recording

The EEG was recorded from 30 standard scalp locations (10–20 system), amplified (band-pass filtered at 0.03 Hz–100 Hz), 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 (see [19,27]). Then, epochs that ranged from −500 ms 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. If the total rejection rate exceeded 40%, data of the participant were excluded, which resulted in the exclusion of 3 participants. Across the remaining participants, average segment loss was 16%.

2.4. Procedure

Participants were seated in front of two loudspeakers, and were informed that they would be listening to short stories. They were instructed to listen for comprehension and minimize movement. No additional task demands were imposed.

Two trial lists were used. For the first list, 30 coherent and 30 anomalous stories were pseudorandomly mixed with 80 filler stories such that neither coherent nor anomalous trials occurred more than two times consecutively and trials of each type were matched on average list position. The second list was derived from the first by replacing all coherent trials by their anomalous counterparts and vice versa. The total of 140 stories was divided in seven blocks, separated by a pause. Each trial was separated from the next by a 5-s silence and

was preceded by a short warning tone. Total time-on-task was approximately 1 h.

3. Results

Fig. 1 displays the grand average waveforms elicited by the coherent animate ("tourist") and anomalous inanimate ("suitcase") critical words at 19 electrode locations. The corresponding difference waveforms, which quantify the net effects of coherence, are displayed in Fig. 2. Two striking results are visible from these figures: In the first place, the absence of a standard N400 effect in the normal 300–500 ms latency range. Secondly, the presence of a large positive deflection emerging around 500–600 ms, with a peak latency within the 900–1100 ms window and a centro-parietal distribution. Using mean amplitude in seven consecutive 200-ms latency windows ranging from 100 to 1500 ms, the overall Coherence (2) × Electrode (30) analysis of variance (ANOVA) revealed main effects of Coherence in the 700–900, 900–1100, and 1100–1300 ms window, but no main effects of Coherence in the 100–300, 300–500, 500–700, or 1300–1500 ms window (Table 2).

Contrary to our expectations, we did not obtain a regular word-elicited N400 effect, an effect that (with spoken language input) emerges at about 150–250 ms after acoustic word onset and is at its maximum in the 400–500 ms latency range. For reasons discussed below, we take the absence of a standard N400 effect as evidence that participants did not immediately notice the replacement of the protagonist by the inanimate object. The large differential effect that started to emerge at about 500–600 ms revealed that listeners did detect the anomaly somewhat later.[1] Note that the onset of this late effect approximates the mean word offset of the impostor words. To examine whether the effect was perhaps triggered by the acoustic offset of our critical words, or by the onset of the next word, we recomputed ERPs for epochs that were time-locked to these two potentially relevant linguistic events. Fig. 3 displays the two associated difference waveforms at P7, Pz, and P8, together with the original difference waveform time-locked to critical word onset (as in Fig. 2). If critical word offset or onset of the next word effectively triggered the late effect observed in Figs. 1 and 2, time-locking to one of these alternative critical events might well yield a larger and more focused ERP effect. However, as can be seen in Fig. 3, time-locking to critical word offset did not yield a sharper and larger differential effect, providing us with little evidence that word offset is by itself critically involved in generating the effect. Additional

---

[1] Because ERPs are average brain potentials, we cannot claim that every participant detected the anomaly in every trial. However, our data do indicate that anomalies were detected in a sufficient number of trials by a sufficient number of participants to generate a large and significant ERP effect.
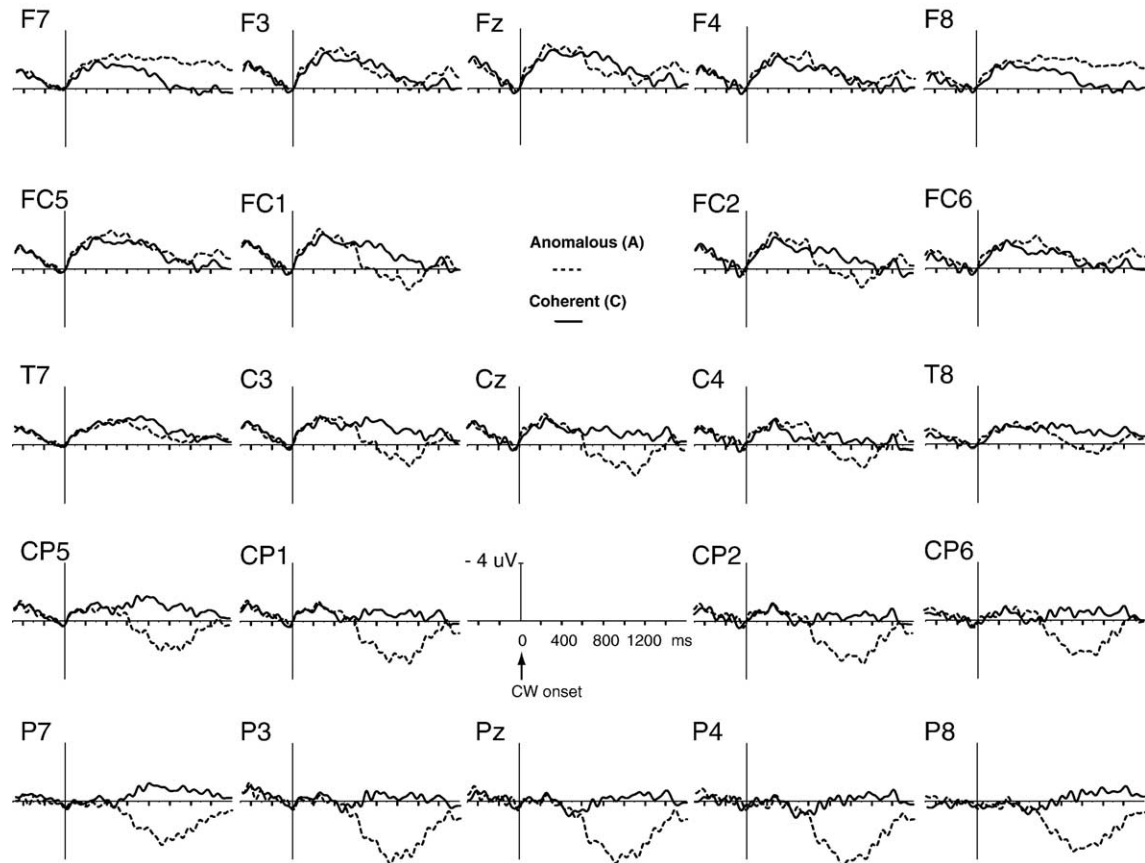
Fig. 1. Grand average ERPs elicited by CWs in the coherent (C) and anomalous (A) condition. In this and all following figures, waveforms are filtered (10 Hz high cut-off, 48 dB/oct) for presentation purpose only, and negativity is plotted upwards.

analysis revealed that the main effect of Coherence did not significantly differ between ERPs time-locked to CW offset and CW onset [mean amplitude in the 200–800 ms window and 600–1200 ms window respectively, $F(1,31) = 0.05$, MSE = 40.42, $P = 0.83$]. Time-locking to the onset of the next word resulted in an evidently smaller ERP effect that was already developing before the 0-ms time-locking point (most prominent at electrode P7). This suggests that the effect at hand is not functionally tied to the onset of the next word. We briefly return to this matter in the discussion.

4. Discussion

The goal of this experiment was to investigate whether conditions that have been reported to induce semantic illusions can interfere with the detection of a very powerful discourse coherence break, as reflected in the magnitude of the N400 effect. To examine this, we unexpectedly replaced one of the central human characters engaged in a conversation in our stories (e.g., a tourist) by a very scenario-relevant but nevertheless inanimate entity (e.g., a suitcase), in a story that was fully coherent up to that point. Because the resulting anomaly deeply affected one of the central characters in the story and hinged on a fundamental

organizing principle of the semantic system (to be alive or not), we expected our listeners to immediately notice the anomaly, in spite of the scenario-relevance, semantic relatedness, and de-accented (i.e., unfocused) acoustic realization of the anomalous word.

Our study yielded two main results, the absence of a regular word-elicited N400 effect and the presence of a differential effect that began to emerge at about 500–600 ms after word onset. We take this pattern of results to suggest that, against our expectations, participants did not immediately notice the replacement of the animate protagonist by an inanimate object. Moreover, we interpret our findings as evidence that participants momentarily suffered from a semantic illusion. In the below, we will unpack each of these claims. After examining these implications of our N400 result, we briefly turn to the late positivity.

4.1. The N400, anomaly detection, and semantic illusions

In the light of what we know about the N400 in language comprehension, the absence of a regular N400 effect in the current study is a very surprising finding. In spoken as well as written language, differences in the degree to which a particular word fits the wider semantic context reliably show up in the amplitude of the N400 elicited by that word (with
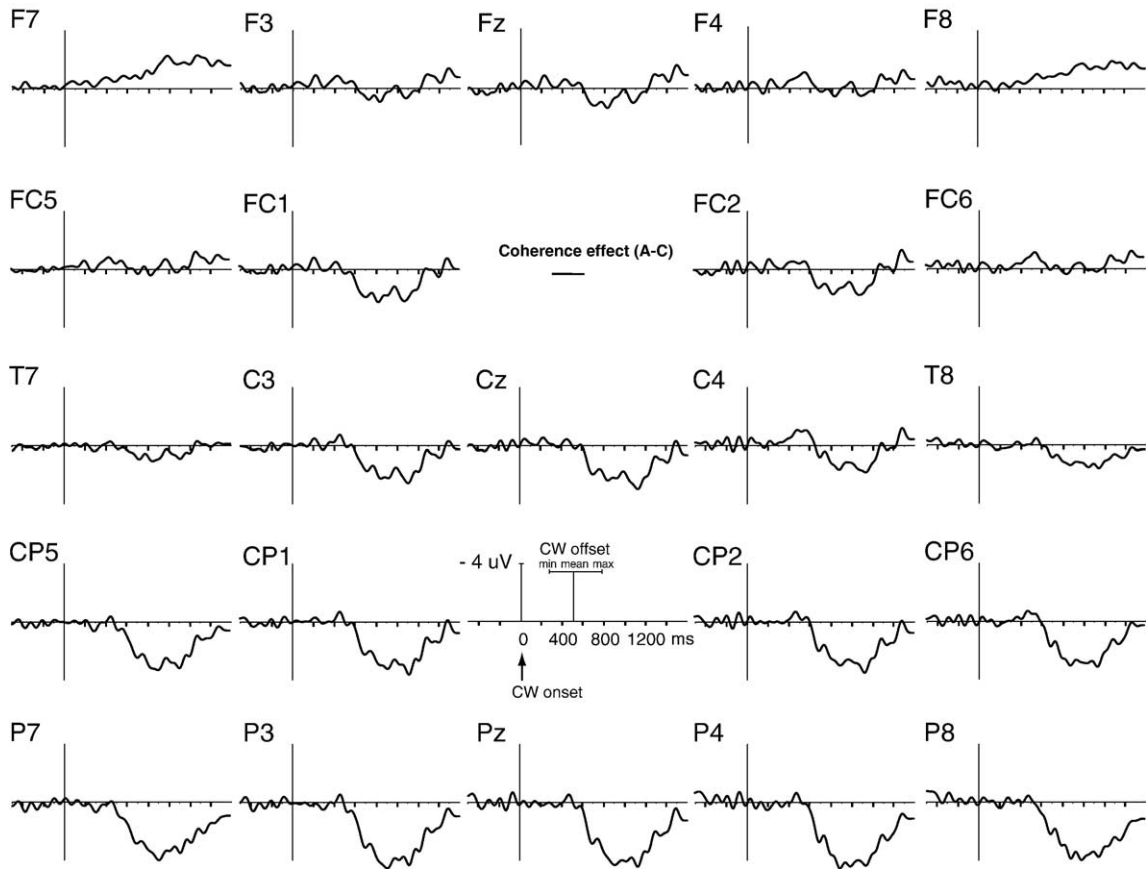
Fig. 2. Coherence effect (anomalous-coherent difference waveform), and minimum, mean, and maximum duration of the critical words.

one interesting exception, to which we return below). A discourse-dependent semantic anomaly, for example, elicits an N400 effect that emerges between 150 and 250 ms after acoustic or visual onset of the critical word, and is invariably large in the 300–500 ms latency range [44,45]. Far more subtle variations in discourse-semantic fit, such as differences in the discourse-based predictability of a word, also reliably elicit N400 effects in this latency range [33,39,46]. Because of the well-established sensitivity of the N400 to even very subtle differences in the ease with which a word can be related to its semantic context, we take the absence of an N400 effect to suggest that in our current experiment, the semantic anomaly in "The woman told the suitcase..." momentarily went undetected.

It is of particular interest to note that verb–object animacy violations of the very type investigated here *did* elicit a standard N400 effect when these sentences were in story-initial position ([30]; see also [52]). In Fig. 4, we display the present ERP results together with those earlier N400 findings, for electrodes Pz, P3, and P4. Note that in either experiment, all we asked our participants to do was to pay attention and listen for comprehension. Our earlier N400 result can thus be taken to confirm that listeners normally detect the animacy violation in "The woman told the suitcase..." immediately, as part of a semantic integration process that is highly sensitive to the degree of fit between current word and prior context. However, when the same critical anomalies are embedded in a wider discourse such that they also disrupt

Table 2
F values, *P* values and mean square errors (MSE) for main effects of Coherence time-locked to critical word onset

| | Latency range from acoustic word onset (ms) | | | | | | |
	100–300	300–500	500–700	700–900	900–1100	1100–1300	1300–1500
F	0.02	0.89	0.95	14.41	10.69	4.64	0.03
MSE	6.96	18.91	36.18	53.42	70.03	82.72	73.34
P	0.905	0.354	0.337	0.001**	0.003**	0.039*	0.876

Note. For all *F* tests, numerator $df = 1$, denominator $df = 31$.
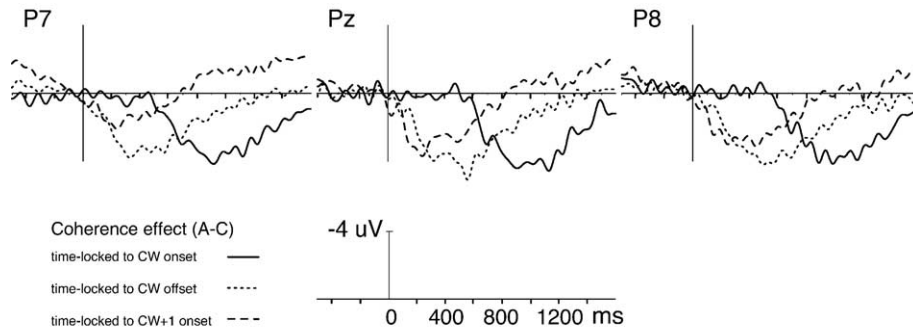 * $P < 0.05$.
** $P < 0.005$.

Fig. 3. Coherence effect (A–C) at P7, Pz, and P8 for epochs time-locked to critical word onset (CW onset), critical word offset (CW offset), and onset of the next word (CW + 1 onset).

discourse-level coherence, and as such effectively present a *twofold* semantic anomaly, they nevertheless momentarily escape attention. Apparently, there are nontrivial conditions under which human story characters can suddenly turn to inanimate objects without the listener noticing this immediately, i.e., at the time at which listeners routinely relate the meaning of an unfolding word to its wider interpretive context.

Because of what is known about the N400 in language comprehension, we believe that the present findings not only demonstrate a temporary anomaly detection failure, but actually also provide evidence for a temporary semantic illusion, i.e., for an apparently coherent but factually incorrect interpretation of the input. The critical thing to note is that, as discussed before, the N400 does not reflect a simple anomaly detection process, but instead reflects a *routine semantic analysis process* that relates the meaning of *every* unfolding word to its wider interpretive context (see [6,23,26] for reviews). In a sentence like "Jenny put the sweet in her...", for instance, the somewhat less expected but perfectly coherent word "pocket" elicits an N400 effect relative to the more expected word "mouth" (e.g., [16,25,39,46]; see [33] for the same phenomenon in connected discourse). Moreover, with spoken language, the routine semantic integration process indexed by the

N400 is known to spring into action after only two or three phonemes of the unfolding word have been heard [45], well before listeners know what the exact word is going to be [47,51], and even well before they can know whether it is a noun, an adjective, or some other part of speech [48].

The established sensitivity of the N400 to subtle modulations of this very early routine semantic analysis process renders several alternative accounts for our findings unlikely. For example, it rules out the possibility that semantic integration is so difficult that *no* interpretation can be arrived at for the new word. If slightly less expected words already elicit a larger N400, a devastatingly problematic word should – if detected as such – certainly do so. It also rules out alternative accounts that focus on the fact that the unfolding word, although problematic as a head noun ('suitcase'), might also be part of a semantically coherent prenominal modifier (e.g., 'suitcase-carrying tourist') or compound noun (e.g., 'suitcase-holder'). The most obvious problem with the latter is that in the language used (Dutch), all such continuations are rather unusual, 'marked' constructions, and are as such far less expected than the straightforward continuation used as baseline condition (e.g., "tourist"). Moreover, even if such less expected continuations might alert the listener to interesting coherent alternative continuations, they should – if detected as such
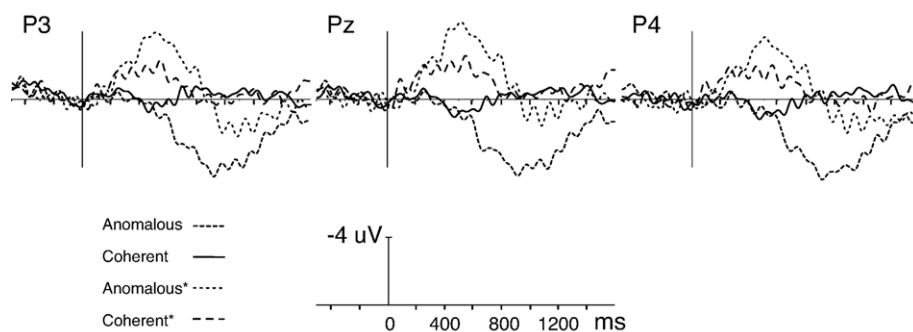

Fig. 4. Overlay of grand average ERPs at electrodes Pz, P3, and P4 elicited by CWs in the coherent (C) and anomalous (A) condition in the present study and in an earlier study ([30], indexed by asterisks) in which we used the exact same set of critical words in similar, but story-initial sentences.

– nevertheless elicit the N400 effect that is reliably elicited by other low-cloze coherent continuations.

The same argument can be leveled against an alternative account in terms of metonymy or reference transfer. Analogous to waitresses referring to an impatient client who ordered a ham sandwich awhile ago as "the ham sandwich is getting restless" [9,31], speakers might in our example decide, for stylistic or other reasons, to refer to the tourist by means of "the suitcase". However, such referential devices are marked, and derive much of their attractiveness as stylistic or humoristic device from the very fact that they are less expected.

In all, and in line with various off-line demonstrations of semantic illusions, we believe that the absence of a regular word-elicited N400 effect at our anomalous words (e.g., "suitcase") should be taken to indicate that our listeners briefly fell prey to a semantic illusion, i.e., that for a few hundred milliseconds, our listeners did have something else in mind, something as unproblematic as the coherent continuation word ("tourist"). In fact, the most parsimonious interpretation of our N400 findings is that our listeners momentarily proceeded as if they had heard the latter word.

Before we can confidently accept the above, we need to examine whether any other aspects of our experiment, other than the conditions that are known to induce semantic illusions, might account for the present pattern of results. First, what about the effects of repetition? Our impostor words were not only semantically similar to the expected words, but had also been repeated several times throughout the preceding discourse. However, although repetition is known to attenuate the N400 amplitude [23,50], it cannot account for the complete absence of an N400 effect in our study (nor for the presence of a later differential effect). The reason is that despite all its repetitions, the anomalous word is still a highly unexpected continuation at that point of the discourse, and should – when detected as such – thus elicit a larger N400.

Second, the absence of an N400 effect can also not be explained by arguing that our participants might have 'tuned out' during the experiment, e.g., because of fatigue or unrealistic materials. As for the latter, we note that the proportion of unrealistic materials in this study is comparable to that in studies in which we did obtain large N400 effects (e.g., [30]). Furthermore, a split-half analysis yielded no differences between coherence effects in the first and second half of the experiment. In addition, if the experiment was such that semantic analysis would be discouraged, anomalous continuations should also not have elicited a differential later ERP effect.

Finally, the absence of an N400 effect is not likely to be the result of active strategic anticipation. As revealed by structured post-session interviews, participants had in the end noticed that some of our stories were rather odd. Of course, it is not inconceivable that at least some (perhaps even all) of the participants came to realize this during the experiment, and may as such have been alerted to the possibility of more upcoming anomalies. However, in the present materials, there is no way to spot the anomalies other than *via regular language comprehension*. The implication is that even if participants were on the lookout for such anomalies (which we are hesitant to believe), the latter should still elicit an N400 effect. Importantly, note that semantic illusion studies have actually shown that detection failures occur with a fairly high frequency even when subjects are explicitly instructed to watch out for anomalies [10]. Thus, even if some of our participants might have been on the lookout, they can still be lured into a temporary semantic illusion.

As it appears, we created a unique combination of conditions that generated a semantic illusion, and as such delayed the detection of a severe anomaly. What might those conditions be? We speculate that besides scenario-relevancy, semantic relatedness, and de-accentuation of the impostor word, a key lies in the anticipation of the correct word in our narratives. There is growing evidence that listeners can use their knowledge of the discourse context to anticipate specific words ([33,42,46]; see also [53]). Other evidence suggests that listeners can begin to look for plausible arguments of a verb right upon encountering the verb itself [2,30,37]. At the critical point in our stories, such as at "Next, the woman told the…" in our example item, these two anticipatory mechanisms may well conspire to strongly suggest both a particular discourse entity that might serve as the verb's argument (the tourist engaged in the conversation), and a particular word ("tourist"). In other words, because discourse constraints at that moment are very powerful, the male protagonist is consequently anticipated as the upcoming argument of the verb, and the language processor essentially takes up an early referential commitment. Altmann [1] persuasively argues that such anticipation need not be particularly odd or effortful, but can instead naturally flow from what linguistic meaning essentially is: knowing when it is appropriate to use certain words or constructions.

How can such anticipation lead to a semantic illusion? To the degree that strong predictions are being made, the initial analysis of what actually comes in might change from signal-driven full analysis to expectation-driven partial matching, or 'verification'. Furthermore, if the unexpected impostor word is highly scenario-relevant, semantically associated with the correct word, and not prosodically accented in the way that an unexpected or newsworthy turn of events would normally be, this partial matching may well temporarily yield a global, good enough fit [4]. Thus, in all, we speculate that in the present study, our severe anomalies are not immediately detected because of strong expectations combined with actual input that has a superficially good fit.

4.2. The late positivity

Of course, if our account is correct, the question remains why the impostor word is detected at all, in our case within some 500–600 ms. The presence of this

differential ERP effect is vital to our interpretation, because it shows that our participants did process for comprehension and did detect the anomaly, albeit somewhat later than expected. Unfortunately, our ERP results provide no specific cues as to what exactly triggered the late positivity. The analyses in which we recomputed ERPs relative to potentially relevant later aspects of the unfolding speech only revealed that it was not triggered by the onset of the next word, nor specifically functionally tied to the acoustic offset of the impostor word.

Taking a constraint-based perspective [15,41], one could argue that in the face of powerful discourse-level and prosodic constraints, it simply takes more time than usual for lexical constraints to build up and overrule the illusory interpretation, but the system does get there in the end. The most natural prediction that would fall out of this account, however, would be that of a delayed (or possibly wider) N400 effect, and additional assumptions are required to explain the qualitatively different ERP effect that we observed. Another possible explanation for our findings takes up a distinction that has also been made in more general accounts of language comprehension that distinguish between a 'quick-and-dirty' heuristics-based initial interpretation and a delayed, more careful interpretive process (e.g., [13,40]). According to such an account, the temporary semantic illusion arises from (or goes undetected because of) the initial use of coarse semantic heuristics, but is soon thereafter unmasked as an incorrect semantic analysis. With respect to the present findings, such an interpretation would suggest that the absence of an N400 reflects the use of these semantic heuristics, which results in a temporary underspecified, and erroneous interpretation, whereas the later positivity would be indicative of a delayed and more careful interpretive process that subsequently overrides this incomplete, illusory interpretation.

It is interesting to note that the combined absence of an expected N400 effect and presence of a later positivity has also been reported in studies on 'semantic reversal anomalies', anomalies that can arise when the arguments of a verb are exchanged (e.g., "The cat that fled from the mice ran through the room" [21]), or more generally, are not in their canonical position (e.g., "For breakfast the eggs would only eat toast and jam" [22]). When the word at which such sentences are rendered semantically anomalous has a strong lexico-semantic fit (e.g., "eat" in the above example), it does not elicit an N400 effect, but instead elicits a late positivity [18,21,22,49]. Because the polarity, timing, and scalp distribution of this positivity strongly resembles that of the syntax-related P600 effect (see [32] for review), it has in all abovementioned studies been taken to be an instance of the latter. The functional account of the P600 effect that is assumed by these authors varies from thematic or syntactic reanalysis in order to obtain a plausible sentence [18,22], to a monitoring process that checks upon the veridicality of ones sentence perception [21,49].

Although there is no logical necessity to interpret the late positivity observed in our experiment as an instance of the P600 effect (because of the inverse problem), we similarly cannot rule out that this is the correct interpretation. This would be of interest, because, even though the animacy violation at a word like "suitcase" hinges on selectional restrictions of the verb, and would thus by some be considered 'close to the syntax' [8], having a conversation with a suitcase is also semantically highly implausible. Moreover, in our materials, the unexpected replacement of a discourse protagonist by an entirely different entity is also a deep violation of discourse-semantic coherence. One potentially viable way to resolve the apparent conflict in obtaining a P600 effect to a – not initially detected – semantic anomaly is to assume that, although the immediate *symptom* that signals a comprehension problem might be semantic, the system might in these cases 'put the blame on syntax' (e.g., consider the possibility of an incorrect phrase ordering). Alternatively, if the P600 more generally indexes a monitoring process that checks upon the veridicality of ones sentence perception, as has been recently proposed [21,49], the conflict would also be resolved.

In the absence of compelling evidence, we refrain from taking a strong position on these issues.[2] Instead, we note that our findings and those from semantic reversal anomalies could be taken to provide converging evidence for the same underlying *temporary semantic illusion* phenomenon, namely, that of a first, superficial, and incorrect semantic analysis, which is followed up by a second, more elaborate interpretive process. That is, readers or listeners may well momentarily believe that "eat" in "For breakfast the eggs would only eat toast and jam" is fine because of some of the same factors that lured our listeners into momentarily believing that "suitcase" is fine: semantic relatedness, scenario-relevance, no accentuation, and strong expectations. Both Van Herten et al. [49] and Hoeks et al. [18] allude to the possibility that their participants might momentarily suffer from a semantic illusion.[3]

---

[2] It is not very likely that the late positivity is an instance of the P300 family, a domain-general brain response elicited by rare and/or informative events [35]. Although anomalous words were relatively rare in our study, there was no other task than to pay attention and listen for comprehension, a requirement that does not in any way impose the task-based explicit response demands involved in, say, oddball detection. But even if the late positivity *would* be part of the P300 family, our N400-based semantic illusion claims still hold. The reason is that the detection of an anomalous 'oddball' can only be carried out by the language comprehension system. Hence, even if the late positivity would be a P300-type effect associated with the detected oddball, the absence of an N400 effect would need to be interpreted along the same lines as we proposed.

[3] Although we are sympathetic with the account of Hoeks and colleagues [18], we are somewhat hesitant to accept their critical claim that these verbs did not elicit an N400 effect. First, the relevant ERP difference wave (Fig. 3) does reveal a small N400 effect. Furthermore, part of this N400 effect may well be counteracted by overlap from an immediately preceding differential positivity (potentially the consequence of processing difficulties elicited by the preceding auxiliary "has"). Finally, we note that their participants were actively conducting a sentence plausibility judgement task, and potentially initiated the associated processing right at the anomalous sentence-final word.

4.3. Semantic change deafness

Regardless of whether the semantic reversal anomaly results can be taken to testify to the same phenomenon, we believe we have evidence that in language comprehension, even attentive listeners can sometimes momentarily fail to notice a radical change in the nature of a critical story character. This has an interesting parallel to the well-documented change blindness phenomenon in visual perception (e.g., [38]), where observers are sometimes oblivious to important changes in the scene—including (under specific conditions) a change in the person they are dealing with. Because the reported temporary semantic illusion involves not noticing that a central protagonist was exchanged for an inanimate object, our ERP findings thus also suggest a case of short-lived 'semantic change deafness' in language comprehension.

Acknowledgments

References

[1] G.T.M. Altmann, The Ascent of Babel: An Exploration of Language, Mind, and Understanding, Oxford University Press, London, 1997.

[2] G.T.M. Altmann, Thematic role assignment in context, J. Mem. Lang. 41 (1999) 124–145.

[3] G.T.M. Altmann, Y. Kamide, Incremental interpretation at verbs: restricting the domain of subsequent reference, Cognition 73 (1999) 247–264.

[4] S.B. Barton, A.J. Sanford, A case study of anomaly detection: shallow semantic processing and cohesion establishment, Mem. Cogn. 21 (1993) 477–487.

[5] S. Bredart, K. Modolo, Moses strikes again: focalization effect on a semantic illusion, Acta Psychol. 67 (1988) 135–144.

[6] C.M. Brown, P. Hagoort, M. Kutas, Postlexical integration processes in language comprehension: evidence from brain-imaging research, in: M.S. Gazzaniga (Ed.), The New Cognitive Neurosciences, The MIT Press, Cambridge, MA, 2000, pp. 881–895.

[7] A. Caramazza, J.R. Shelton, Domain-specific knowledge systems in the brain: the animate–inanimate distinction, J. Cogn. Neurosci. 10 (1998) 1–34.

[8] N. Chomsky, Aspects of the Theory of Syntax, The MIT Press, Cambridge, MA, 1965.

[9] H.H. Clark, Arenas of Language Use, University of Chicago Press, Chicago, 1992.

[10] T.D. Erickson, M.E. Mattson, From words to meaning: a semantic illusion, J. Verbal Learn. Verbal Behav. 20 (1981) 540–551.

[11] K.D. Federmeier, D.B. McLennan, E. De Ochoa, M. Kutas, The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study, Psychophysiology 39 (2002) 133–146.

[12] F. Ferreira, K.G.D. Bailey, V. Ferraro, Good-enough representa-

[13] S. Garrod, A. Sanford, Incrementality in discourse understanding, in: S.R. Goldman, H. van Oostendorp (Eds.), The Construction of Mental Representations during Reading, Lawrence Erlbaum Associates, Mahwah, 1999, pp. 3–27.

[14] S.A. Gelman, J.E. Opfer, Development of the animate–inanimate distinction, in: U. Goswami (Ed.), London Medical School Blackwell Handbook of Childhood Cognitive Development, Blackwell Publisher, Malden, 2002, pp. 151–166.

[15] E. Gibson, N.J. Pearlmutter, Constraints on sentence comprehension, Trends Cogn. Sci. 2 (7) (1998) 262–268.

[16] P. Hagoort, C. Brown, Brain responses to lexical ambiguity resolution and parsing, in: C. Clifton, L. Frazier, K. Rayner (Eds.), Perspectives on Sentence Processing, Erlbaum, Hillsdale, 1994, pp. 45–80.

[17] D.J. Hess, D.J. Foss, P. Carroll, Effects of global and local context on lexical processing during language comprehension, J. Exp. Psychol. Gen. 124 (1995) 62–82.

[18] J.C.J. Hoeks, L.A. Stowe, G. Doedens, Seeing words in context: the interaction of lexical and sentence level information during reading, Cogn. Brain Res. 18 (2004) 59–73.

[19] T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T.J. Sejnowski, Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects, Clin. Neurophysiol. 111 (2000) 1745–1758.

[20] D. Kahneman, A. Tverski, Choices, Values, and Frames, Cambridge University Press, Cambridge, 2000.

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

[22] G.R. Kuperberg, T. Sitnikova, D. Caplan, P.J. Holcomb, Electrophysiological distinctions in processing conceptual relationships within simple sentences, Cogn. Brain Res. 17 (2003) 117–129.

[23] M. Kutas, K.D. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, Trends Cogn. Sci. 4 (12) (2000) 463–470.

[24] M. Kutas, S.A. Hillyard, Reading senseless sentences: brain potentials reflect semantic incongruity, Science 207 (1980) 203–205.

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

[26] M. Kutas, C.K. Van Petten, Psycholinguistics electrified: event-related brain potential investigations, in: M.A. Gernsbacher (Ed.), Handbook of Psycholinguistics, Academic Press, San Diego, 1994, pp. 83–143.

[27] S. Makeig, T.-P. Jung, D. Ghahremani, A.J. Bell, T.J. Sejnowski, Blind separation of event-related responses into independent components, Proc. Nat. Acad. Sci. U. S. A. 94 (1997) 10979–10984.

[28] W. Marslen-Wilson, Sentence perception as an interactive parallel process, Science 189 (1975) 226–228.

[29] W. Marslen-Wilson, L.K. Tyler, The temporal structure of spoken language understanding, Cognition 8 (1) (1980) 1–71.

[30] M.S. Nieuwland, J.J.A. Van Berkum, Discourse context can completely overrule lexical–semantic violations: evidence from the N400 (in preparation).

[31] G. Nunberg, The Pragmatics of Reference, Indiana University Linguistics Club, Bloomington, 1978.

[32] L. Osterhout, P. Hagoort, 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, Lang. Cogn. Processes 14 (1999) 1–14.

[33] M. Otten, J.J.A. Van Berkum, Discourse based lexical anticipation during language processing: prediction or priming? (in preparation).

[34] S. Pinker, How the Mind Works, W W Norton and Co Inc, New York, 1997.

[35] W.S. Pritchard, Psychophysiology of P300, Psychol. Bull. 89 (1981) 506–540.

[36] A.J. Sanford, P. Sturt, Depth of processing in language compre-

hension: not noticing the evidence, Trends Cogn. Sci. 6 (2002) 382–386.

[37] J.C. Sedivy, M.K. Tanenhaus, C.G. Chambers, G.N. Carlson, Achieving incremental semantic interpretation through contextual representation, Cognition 71 (1999) 109–147.

[38] D.J. Simons, D.T. Levin, Change blindness, Trends Cogn. Sci. 1 (7) (2000) 261–267.

[39] M. St. George, S. Mannes, J.E. Hoffman, Global semantic expectancy and language comprehension, J. Cogn. Neurosci. 6 (1994) 70–83.

[40] D.J. Townsend, T.G. Bever, Sentence Comprehension: The Integration of Habits and Rules, MIT Press, Cambridge, 2001.

[41] M.K. Tanenhaus, C. Trueswell, Sentence comprehension, in: J.L. Miller, P.D. Eimas (Eds.), Speech, Language, and Communication, Academic Press, San Diego, 1995, pp. 217–262.

[42] J.J.A. Van Berkum, Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? in: M. Carreiras, C. Clifton Jr. (Eds.), The On-Line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond, Psychology Press, New York, 2004, pp. 229–270.

[43] J.J.A. Van Berkum, C.M. Brown, P. Hagoort, Early referential context effects in sentence processing: evidence from event-related brain potentials, J. Mem. Lang. 41 (1999) 147–182.

[44] J.J.A. Van Berkum, P. Hagoort, C.M. Brown, Semantic integration in sentences and discourse: evidence from the N400, J. Cogn. Neurosci. 11 (1999) 657–671.

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

[46] J.J.A. Van Berkum, C.M. Brown, P. Zwitserlood, V. Kooijman, P. Hagoort, Anticipating upcoming words in discourse: evidence from ERPs and reading times. In press, J. Exper. Psychol., Learn., Mem., Cogn. 31 (3) (2004).

[47] D. Van den Brink, C.M. Brown, P. Hagoort, Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects, J. Cogn. Neurosci. 13 (2001) 967–985.

[48] D. Van den Brink, P. Hagoort, The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs, J. Cogn. Neurosci. 16 (2004) 1068–1084.

[49] M. Van Herten, H.H.J. Kolk, D.J. Chwilla, An ERP study of P600 effects elicited by semantic reversal anomalies, Cogn. Brain Res. 22 (2005) 241–255.

[50] C. Van Petten, M. Kutas, R. Kluender, M. Mitchiner, H. McIsaac, Fractionating the word repetition effect with event-related potentials, J. Cogn. Neurosci. 3 (2) (1991) 131–150.

[51] C. Van Petten, S. Coulson, S. Rubin, E. Plante, M. Parks, Time course of word identification and semantic integration in spoken language, J. Exper. Psychol., Learn., Mem., Cogn. 25 (1999) 394–417.

[52] J. Weckerly, M. Kutas, An electrophysiological analysis of animacy effects in the processing of objective relative sentences, Psychophysiology 36 (1999) 559–570.

[53] N.Y.Y. Wicha, E.M. Moreno, M. Kutas, Expecting gender: an event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish, Cortex 39 (2003) 483–508.

[54] R.A. Zwaan, G.A. Radvansky, Situation models in language comprehension and memory, Psychol. Bull. 123 (1998) 162–185.