

# Beyond the sentence given

Peter Hagoort<sup>1,2,\*</sup> and Jos van Berkum<sup>1,2,3</sup>

<sup>1</sup>*EC. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen,  
PO Box 9101, 6500 HB Nijmegen, The Netherlands*

<sup>2</sup>*Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands*

<sup>3</sup>*Department of Psychology, University of Amsterdam, Roeterstraat 15,  
1018 WB Amsterdam, The Netherlands*

A central and influential idea among researchers of language is that our language faculty is organized according to Fregean compositionality, which states that the meaning of an utterance is a function of the meaning of its parts and of the syntactic rules by which these parts are combined. Since the domain of syntactic rules is the sentence, the implication of this idea is that language interpretation takes place in a two-step fashion. First, the meaning of a sentence is computed. In a second step, the sentence meaning is integrated with information from prior discourse, world knowledge, information about the speaker and semantic information from extra-linguistic domains such as co-speech gestures or the visual world. Here, we present results from recordings of event-related brain potentials that are inconsistent with this classical two-step model of language interpretation. Our data support a one-step model in which knowledge about the context and the world, concomitant information from other modalities, and the speaker are brought to bear immediately, by the same fast-acting brain system that combines the meanings of individual words into a message-level representation. Underlying the one-step model is the immediacy assumption, according to which all available information will immediately be used to co-determine the interpretation of the speaker's message. Functional magnetic resonance imaging data that we collected indicate that Broca's area plays an important role in semantic unification. Language comprehension involves the rapid incorporation of information in a 'single unification space', coming from a broader range of cognitive domains than presupposed in the standard two-step model of interpretation.

**Keywords:** language; event-related brain potentials; functional magnetic resonance imaging; discourse; semantic unification; Broca's area

## 1. INTRODUCTION

As a result of the Chomskyan revolution in linguistics (Chomsky 1957), theories about human language comprehension often assume that the sentence is not only the core unit of syntactic analysis, but also the core unit of language interpretation. The assumption follows from the fact that the sentence is the domain of syntactic analysis coupled with two dominant ideas in mainstream generative grammar: (i) the truly relevant combinatorics of language are coded in the syntax and (ii) the semantic interpretation of an expression is derived from its syntactic structure. The latter idea is what Culicover & Jackendoff (2006) have recently referred to as Fregean compositionality, the claim that the overall meaning of an utterance is a function of the meaning of its parts and of the syntactic rules by which they are combined.

The implication of this idea is that language interpretation takes place in a two-step fashion. First, the context-free meaning of a sentence is computed by combining fixed word meanings in ways specified by the syntax. In a second step, the sentence meaning is integrated with information from prior discourse,

world knowledge, information about the speaker and semantic information from extra-linguistic domains such as co-speech gestures or the visual world. The latter step is needed because interpretation is clearly shaped by factors beyond the sentence given. That is, listeners interpret language not only by combining stored word meanings in accordance with the grammar, but also by taking into consideration their knowledge about the speaker (Clark 1996), their knowledge of the world (Jackendoff 2003) and the available information from the other input modalities (Tanenhaus *et al.* 1995).

There is widespread agreement that such additional 'contextual' factors help to fix the final interpretation of a sentence. However, there is disagreement over whether such factors can also immediately co-determine the initial interpretation of sentence-level expressions. The standard two-step model of interpretation prohibits such immediate contextualization of meaning (e.g. Grice 1975; Fodor 1983; Sperber & Wilson 1995; Cutler & Clifton 1999; Lattner & Friederici 2003). For instance, in their blueprint of the listener, Cutler & Clifton (1999) assume that, based on syntactic analysis and thematic processing, utterance interpretation takes place first, in a next processing step integration into a discourse model follows. Along similar lines, Lattner & Friederici (2003)

\* Author for correspondence (peter.hagoort@fcdonders.ru.nl).

One contribution of 14 to a Discussion Meeting Issue 'Mental processes in the human brain'.

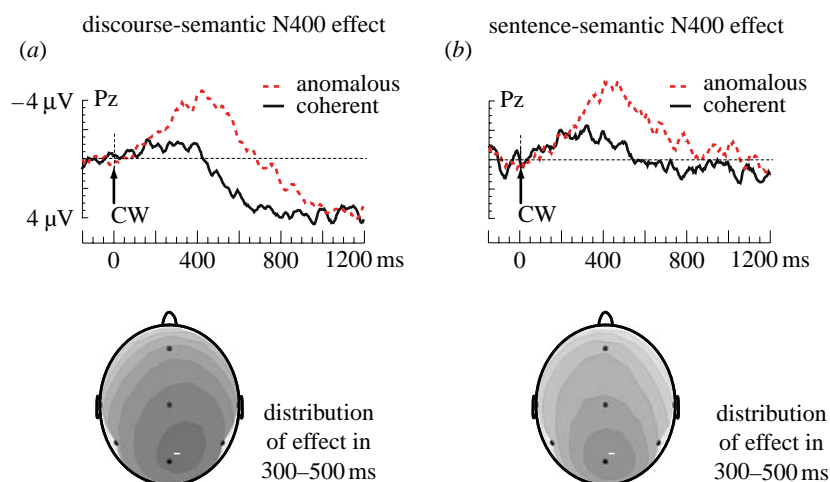


Figure 1. N400 effects triggered by (a) discourse-related and (b) sentence-related anomalies. Waveforms are presented for a representative electrode site (Pz). The latencies of the N400 effect in discourse and sentence contexts (both onset and peak latencies) are the same (after Van Berkum *et al.* 2003).

recently argued that mismatches between spoken message and speaker are detected relatively late, in slow pragmatic computations that are different from the rapid semantic computations in which word meanings are combined. Adherents of a one-step model of language interpretation, in contrast, take the immediacy assumption as their starting point (cf. Just & Carpenter 1980), i.e. the idea that every source of information that constrains the interpretation of an utterance (syntax, prosody, word-level semantics, prior discourse, world knowledge, knowledge about the speaker, gestures, etc.) can in principle do so immediately (e.g. Crain & Steedman 1985; Garrod & Sanford 1994; MacDonald *et al.* 1994; Tanenhaus & Trueswell 1995; Clark 1996; Altmann 1997; Van Berkum *et al.* 1999; Jackendoff 2002; Zwaan 2004).

In our contribution, we review the results of a number of studies that aimed to determine the processing principles of language understanding beyond the sentence level and that are directly relevant to the issue of one- versus two-step language interpretation. We looked at the influence of discourse, world knowledge and co-speech gestures on the integration of lexical information into a coherent mental model of what is being talked about ('situation model'; Zwaan & Radvansky 1998). In most of these studies, we made use of event-related brain potentials (ERPs), an average measure of electroencephalogram (EEG) activity associated with particular critical events. Because ERPs provide a direct and qualitative informative record of neuronal activity, with almost 0 ms delay, they allow one to keep track of the various processes in language comprehension with high temporal resolution. For several of the studies discussed below, we also briefly report functional magnetic resonance imaging (fMRI) data collected with the same experimental design to identify crucial cortical contributions to language interpretation.

## 2. THE DOMAIN OF SEMANTIC UNIFICATION: SENTENCE VERSUS DISCOURSE

To investigate the different claims of the one-step and the two-step models empirically, we first conducted an ERP

study aiming to unravel how and when the language comprehension system relates an incoming spoken word to semantic representations of the unfolding local sentence and the wider discourse (Van Berkum *et al.* 2003). For this and most of the other studies discussed here, we exploited the characteristics of the so-called N400 component in the ERP waveform. Kutas & Hillyard (1980) were the first to observe this negative-going potential with an onset at approximately 250 ms and a peak at approximately 400 ms (hence the N400), whose amplitude was increased when the semantics of the eliciting word (i.e. *socks*) mismatched with the semantics of the sentence context, as in *He spread his warm bread with socks*.

Since its original discovery in 1980, much has been learned about the processing nature of the N400 (for extensive overviews, see Kutas & Van Petten 1994; Osterhout & Holcomb 1995; Kutas *et al.* 2006; Osterhout *et al.* in press). In particular, as Hagoort & Brown (1994) and many others have observed, the N400 effect does not depend on a semantic violation. For example, subtle differences in semantic expectancy, as between *mouth* and *pocket* in the sentence context 'Jenny put the sweet in her *mouth/pocket* after the lesson', can also modulate the N400 amplitude (Hagoort & Brown 1994). Specifically, as the degree of semantic fit between a word and its context increases, the amplitude of the N400 goes down. Owing to such subtle modulations, the word-elicited N400 is generally viewed as reflecting the processes that integrate the meaning of a word into the overall meaning representation constructed for the preceding language input (Osterhout & Holcomb 1992; Brown & Hagoort 1993).

In our discourse experiment (Van Berkum *et al.* 2003; see Van Berkum *et al.* (1999) for a written-language variant), listeners heard short stories of which the last sentence sometimes contained a critical word that was semantically anomalous with respect to the wider discourse (e.g. *Jane told the brother that he was exceptionally slow* in a discourse context where he had in fact been very quick). Relative to a discourse-coherent counterpart (e.g. *quick*), these discourse-anomalous words (*slow* in the example sentence) elicited a large

N400 effect (i.e. a negative shift in the ERP that began at approximately 150–200 ms after spoken word onset and peaked around 400 ms; [figure 1a](#)).

Next to the discourse-related anomalies, standard sentence-semantic anomaly effects were elicited under comparable experimental conditions ([figure 1b](#)). The ERP effects elicited by both types of anomalies were highly similar. Relative to their coherent counterparts, discourse- and sentence-anomalous words elicited an N400 effect with an identical time course and scalp topography ([figure 1](#)). The similarity of these effects, particularly in polarity and scalp distribution, is compatible with the claim that they reflect the activity of a largely overlapping or identical set of underlying neural generators, indicating similar functional processes. In related studies, we have furthermore found that like sentence-dependent N400 effects, discourse-dependent N400 effects can also be elicited by coherent words that are simply somewhat less expected ([Van Berkum et al. 2005](#); [Otten & Van Berkum in press](#)).

In line with other work (e.g. [St George et al. 1994](#)), our discourse ERP studies provide no indication whatsoever that the language comprehension system is slower in relating a new word to the semantics of the wider discourse than in relating it to local sentence context. Our findings thus do not support the idea that new words are related to the discourse model after they have been evaluated in terms of their contribution to local sentence semantics. Furthermore, the speed with which discourse context affects processing of the current sentence appears to be at odds with estimates of how long it would take to retrieve information about prior discourse from long-term memory. In the material of [Van Berkum and colleagues](#), the relative coherence of a critical word usually hinged on rather subtle information that was implicit in the discourse and required considerable inferencing about the discourse topic and the situation it described. [Kintsch \(Ericsson & Kintsch 1995; Kintsch 1998\)](#) has suggested that during online text comprehension, such subtle discourse information is not immediately available and must be retrieved from memory when needed. This is estimated to take some 300–400 ms at least. However, the results of our experiments suggest that the relevant discourse information can sometimes be brought to bear on local processing within a mere 150 ms after spoken word onset.

As discussed elsewhere ([Van Berkum et al. 1999, 2003](#)), the observed identity of discourse- and sentence-level N400 effects can be accounted for in terms of a processing model that abandons the distinction between sentence- and discourse-level semantic unification. One viable way to do this (in our view) is by invoking the notion of ‘common ground’ ([Stalnaker 1978, Clark 1996](#)). Linguistic analyses have demonstrated that the meaning of utterances cannot be determined without taking into account the knowledge that speaker and listener share and mutually believe they share. This common ground includes a model of the discourse itself (i.e. a situation model as well as a record of the exchange, ‘a discourse record’ or ‘textbase’; [Clark 1996](#)), which is continually updated as the discourse unfolds. If listeners and readers always immediately evaluate new words relative to the

discourse model and the associated information in common ground (i.e. immediately compute ‘contextual meaning’), the identity of the ERP effects generated by sentence- and discourse anomalies has a natural explanation. With a single sentence, the relevant common ground only includes whatever discourse and world knowledge has just been activated by the sentence fragment presented so far. With a sentence presented in discourse context, the relevant common ground will be somewhat richer, now also including information elicited by the specific earlier discourse. But the unification process that integrates incoming words with the relevant common ground should not really care about where the interpretive constraints came from. We suspect that the N400 effects observed by [Van Berkum et al. \(2003\)](#) reflect the activity of this single conceptual unification process.

Of course, this is not to deny the relevance of sentential structure for semantic interpretation. In particular, how the incoming words are related to the discourse model is co-constrained by sentence-level syntactic devices (such as word order, case marking, local phrase structure or agreement) and the associated mapping onto thematic roles. However, this is fully compatible with the claim that there is no separate stage during which word meaning is exclusively evaluated with respect to ‘local sentence meaning’, independent of the discourse context in which that sentence occurs.

The idea that language interpretation involves the immediate mapping of incoming word meanings onto the widest interpretive domain available has also received supported from eye tracking data with readers (e.g. [Hess et al. 1995](#)) and listeners (e.g. [Altmann & Kamide 1999; Hanna et al. 2003](#); see [Trueswell & Tanenhaus \(2005\)](#) for review). However, unlike eye movements, brain potentials provide clear cues to the identity of the processes involved, and therefore allow for stronger inferences about whether or not two sources of information are recruited by the same neuronal system ([Van Berkum 2004](#)). It is due to this feature that ERP data can make a unique contribution to debates about the (non)equivalence of specific processes.

Particularly strong ERP evidence for the immediate integration of lexical-semantic information into a discourse model has recently been provided by [Nieuwland & Van Berkum \(2006\)](#). They had subjects listening to short stories in which the inanimate protagonist was attributed with different animacy characteristics. Here is an example of the materials, with the critical words in italics:

A woman saw a dancing peanut who had a big smile on his face. The peanut was singing about a girl he had just met. And judging from the song, the peanut was totally crazy about her. The woman thought it was really cute to see the peanut singing and dancing like that. The peanut was *salted/in love*, and by the sound of it, this was definitively mutual. He was seeing a little almond.

As can be seen in [figure 2](#), the canonical inanimate predicate (*salted*) for this inanimate object (peanut) elicited a larger N400 than the locally anomalous, but contextually appropriate predicate (i.e. a peanut that is in love).

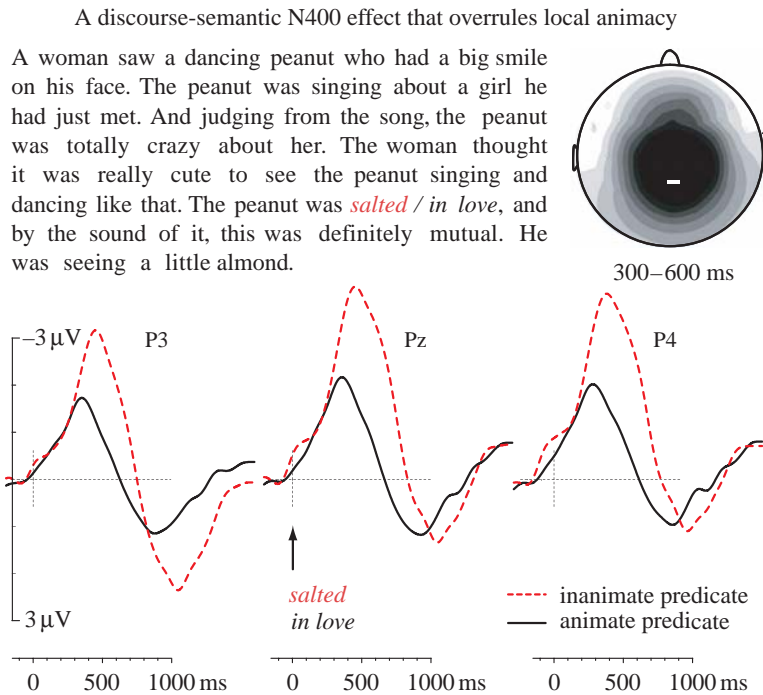


Figure 2. N400 effects triggered by a correct predicate (*salted*) that is, however, contextually disfavoured in comparison to an incorrect predicate (*in love*). Waveforms are presented for representative electrode sites, time locked to the onset of the critical inanimate/animate predicate in the fifth sentence (after Nieuwland & Van Berkum 2006).

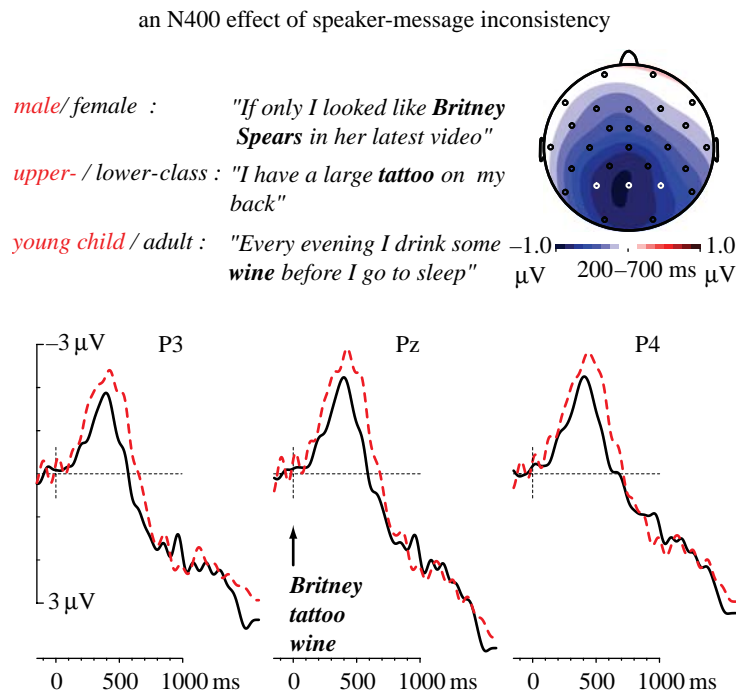


Figure 3. N400 effects triggered by a critical word (in bold) that rendered the spoken sentence inconsistent with voice-based inferences about the speaker. Three representative electrode sites are shown (speaker-inconsistent waveforms are in red) as well as the topographic distribution of the N400 effect (Van Berkum *et al.* submitted).

These results show that discourse context can completely overrule constraints provided by animacy, a feature claimed to be part of the evolutionary hardwired aspects of conceptual knowledge (Caramazza & Shelton 1998), and often mentioned as a prime example of the semantic primitives involved in the computation of context-free sentence meaning (cf. Fregean compositionality). As such, these ERP results provide strong evidence against the standard two-step model of language interpretation.

### 3. KNOWLEDGE OF THE SPEAKER

In interpreting a speaker's utterance, we not only take the preceding utterances into consideration, but also our knowledge of the speaker. For instance, we know that a toddler is unlikely to say 'I studied quantum physics during my holidays', and that it is really odd for a man to say 'I think I am pregnant because I feel sick every morning'. As examples such as these reveal, at some point during language comprehension, people combine the information that is represented in the

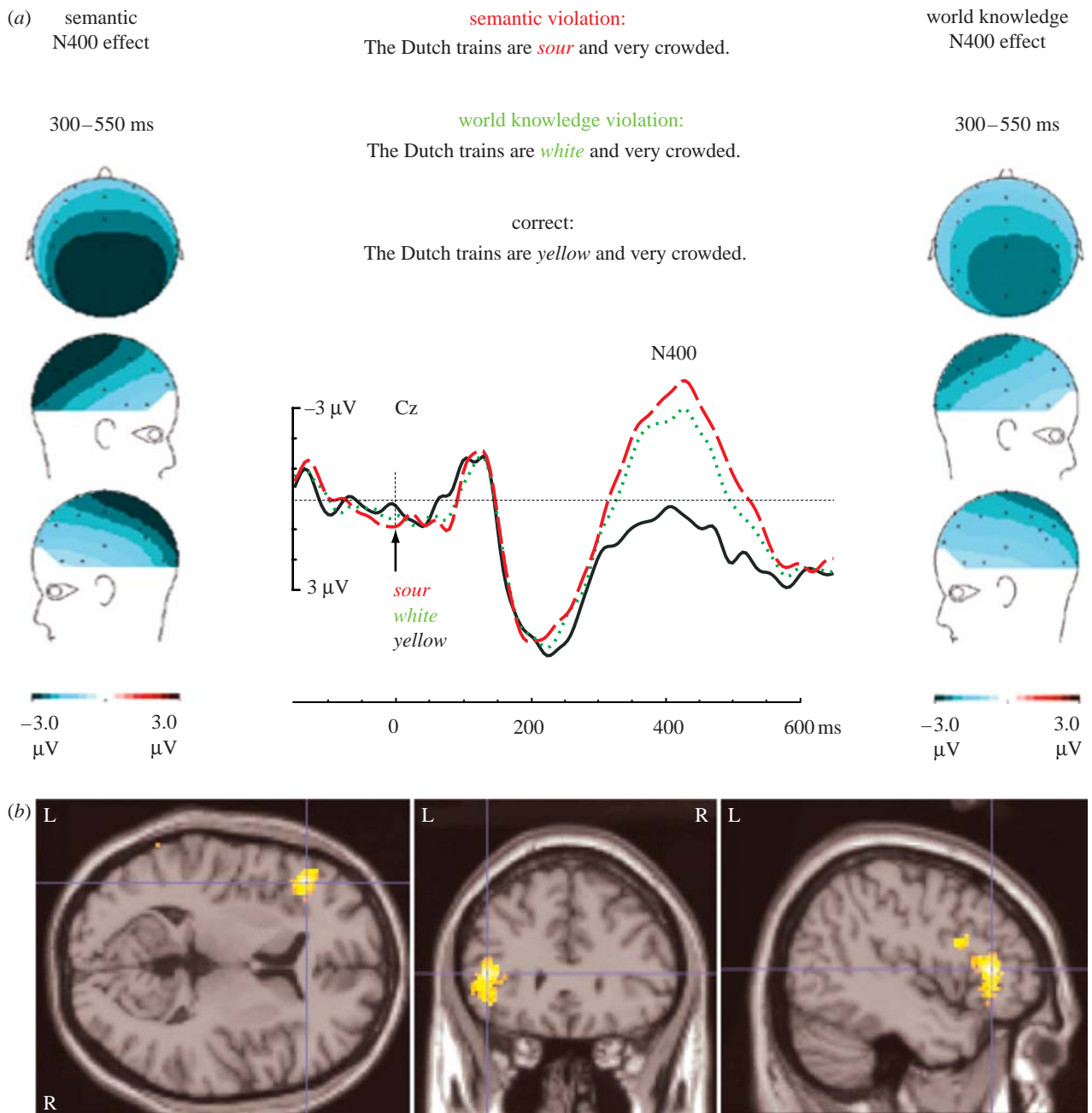


Figure 4. (a) Grand average ERPs for a representative electrode site (Cz) for correct condition (black line), world knowledge violation (green dotted line) and semantic violation (red dashed line). ERPs are time locked to the presentation of the critical words (in italic). Spline-interpolated isovoltage maps display the topographic distributions of the mean differences from 300 to 550 ms between semantic violation and control (left); and between world knowledge violation and control (right). Topographic distributions of the N400 effect are not significantly different between semantic and world knowledge violation ( $p=0.9$ ). (b) The common activation for semantic and world knowledge violations compared with the correct condition based on the results of a minimum-T-field conjunction analysis. Both violations resulted in a single common activation ( $p=0.043$ , corrected) in the LIFG (in, or in the vicinity of, Brodmann's area 45 ( $[x, y, z] = [-44, 30, 8]$ ;  $Z=4.87$ ) and brain area (BA 47) ( $[x, y, z] = [-48, 28, -12]$ ;  $Z=4.15$ ). The cross hair indicates the voxel of maximal activation and has the following coordinates  $[x, y, z] = [-44, 30, 8]$  (left BA 45).

contents of a sentence with the information they have about the speaker. The question again concerns exactly when the pragmatic information about the speaker is having its impact on the unfolding interpretation of the utterance.

In an ERP experiment (Van Berkum *et al.* submitted), people listened to sentences, some of which contained a speaker inconsistency, a specific word at which the message content became at odds with inferences about the speaker's sex, age and social

status, as inferred from the speaker's voice. One example was: 'I have a large *tattoo* on my back' spoken in an upper-class accent. For comparison, other sentences contained a standard semantic anomaly, a specific word whose meaning did not fit the semantic context established by the preceding words, as in 'The Earth revolves around the *trouble* in a year'.

If voice-based inferences about the speaker are recruited by the same early unification process that combines word meanings, then speaker inconsistencies

and semantic anomalies should elicit the same N400 effect (though not necessarily of the same size). But if, as predicted by the two-step model of semantic interpretation, contextual information about the speaker is handled in a distinct second phase of interpretation (cf. Lattner & Friederici 2003), then speaker inconsistencies should elicit a delayed and possibly quite different ERP effect. As can be seen in figure 3, speaker inconsistencies elicited a small but clear N400 effect with a classical posterior maximum. Moreover, its onset latency is the same as for the standard N400 effect. Importantly, reliable effects of speaker inconsistency were already found in the 200–300 ms latency range after word onset. The same latency effects were obtained in this experiment for the straightforward semantic anomalies.

According to our ERP results, the brain integrates message content and speaker information within some 200–300 ms after the acoustic onset of a relevant word. Also, speaker inconsistencies elicited the same type of brain response as semantic anomalies, an N400 effect. That is, voice inferred information about the speaker is taken into account by the same early language interpretation mechanisms that construct ‘sentence-internal’ meaning based on just the words. These findings therefore demonstrate again that linguistic meaning depends on the pragmatics of the communicative situation right from the start. However, by revealing an immediate impact of what listeners infer about the speaker, the present results add a distinctly social dimension to the mechanisms of online language interpretation. What we see is that language users immediately model the speaker to help determine what is being said. This ERP finding converges with linguistic analyses of conversation (Clark 1996) as well as with evidence from eye movements for the rapid use of speaker-related information during comprehension (e.g. Hanna *et al.* 2003; Trueswell & Tanenhaus 2005).

In addition, in an fMRI version of this experiment, we found that the increased unification load of combining incompatible speaker information and message content resulted in increased activation of the left inferior frontal gyrus (LIFG), the area that has been found to be of importance for unification operations in many other neuroimaging studies (cf. Hagoort 2005).

#### 4. WORLD KNOWLEDGE VERSUS SEMANTIC KNOWLEDGE

At least since Frege (1892, see Seuren 1998), theories of meaning make a distinction between the semantics of an expression and its truth-value in relation to our mental representation of the state of affairs in the world (Jackendoff 2002). For instance, the sentence ‘The present Queen of England is divorced’ has a coherent semantic interpretation, but contains a proposition that is false in the light of our knowledge in memory that Her Majesty is married to Prince Phillip. The situation is different for the sentence ‘The favorite palace of the present Queen of England is divorced’. Under default interpretation conditions, this sentence has no coherent semantic interpretation, since the predicate *is divorced* requires an animate argument. This sentence

mismatches with our representation of the world in memory, because the descriptive features of the purported state of affairs are inherently in conflict. The difference between these two sentences points to the distinction that can be made between facts of the world and the words of our language, including their meaning (lexical semantics). In the standard two-step model of interpretation, only the latter type of knowledge feeds into the construction of initial sentence meaning; the integration of pragmatic or world knowledge information would be delayed and handled by a different system (e.g. Sperber & Wilson 1986).

Hagoort *et al.* (2004) performed a combined EEG/MRI study that speaks to this issue. While participants’ brain activity was recorded, they read three versions of sentences such as: ‘The Dutch trains are *yellow/white/sour* and very crowded.’ (the critical words are in italics). It is a well-known fact among Dutch people that Dutch trains are yellow and, therefore, the first version of this sentence is correctly understood as true. However, the linguistic meaning of the alternative colour term *white* applies equally well to trains as the predicate *yellow*. It is world knowledge about trains in Holland that makes the second version of this sentence false. This is different for the third version, where (under standard interpretation conditions) the core semantic features of the predicate *sour* do not fit the semantic features of its argument *trains*. One could thus argue that the third sentence is false or incoherent for semantic-internal reasons: it is our knowledge about the words of our language and their linguistic meaning that poses a problem. If semantic interpretation precedes verification against world knowledge, the effects of the semantic violations should be earlier and might invoke other brain areas than the effects of the world knowledge violations.

Figure 4 presents an overview of the results. As expected, the classic N400 effect was obtained for the semantic violations. For the world knowledge violations, a clear N400 effect was observed as well. Crucially, this effect was identical in onset and peak latency, and very similar in amplitude and topographic distribution to the semantic N400 effect. This finding is strong empirical evidence that lexical-semantic knowledge and general world knowledge are both integrated in the same time-frame during sentence interpretation, starting at approximately 300 ms after word onset. Furthermore, the fMRI data (figure 4b), time locked to the onset of the critical words, revealed a common activation increase in LIFG for both semantic and world knowledge violations, when compared with correct sentences, observed in Brodmann’s areas 45 and 47.

Both word meaning and world knowledge are thus recruited and integrated very rapidly, that is within some 400 ms, during online sentence comprehension. The LIFG, including Broca’s area, seems to be critical both in the computation of meaning and in the verification of linguistic expressions. Although Frege (1892) made an important distinction between the sense of a proposition and relating it to the states of affairs in the world, the processing consequences of a lexical-semantic and world-knowledge problem appear to be immediate and parallel.<sup>1</sup> The results of our world-

knowledge experiments, therefore, provide further evidence against a non-overlapping two-step unification process in which first the meaning of a sentence is determined, and only then its meaning is verified in relation to our knowledge of the world. Semantic interpretation is not separate from its integration with non-linguistic elements of meaning.

## 5. THE INTEGRATION OF CO-SPEECH GESTURES

In ordinary face-to-face conversation, language users not only hear speech but also see the speaker's hand, mouth and body movements. This concurrent visual information often bears on the message conveyed. For example, when talking about drinking a glass of whisky, speakers sometimes perform a concomitant drink gesture (i.e. C shaped hand moved towards the mouth) as they utter the verb 'drink' in their spoken utterance. The listener's brain therefore continuously integrates spoken language information with several streams of visual information, including information from the lips, the eyes and, crucially, semantic information from the hand gestures, that accompany speech (McNeill 1992). Yet, until recently, nothing was known about whether and how listeners integrate the semantic information from co-speech gestures online into the discourse model, and about how this compares to the discourse-model integration of spoken words.

In two recent ERP and fMRI studies (Özyürek *et al.* in press; Willems *et al.* in press), we have begun to address the issue by focusing on iconic gestures that convey information about the shape, size, motion and action characteristics of the events described in the spoken utterance. To determine the nature of the integration of verbal and gestural semantic information, we manipulated the semantic fit of speech (i.e. a critical verb) and/or gesture in relation to the preceding part of the sentence, as well as the semantic relations between the temporally overlapping gesture and speech (table 1).

Movie clips of the iconic gestures were temporally aligned to the critical verbs in the sentences. This manipulation resulted in four conditions (table 1): in the language 'mismatch' condition, the critical verb was harder to fit semantically to the preceding context while the co-occurring gesture matched the sentence context perfectly. In the gesture mismatch condition, the gesture was harder to integrate to previous context, while the critical verb matched the spoken sentence context. In the condition in which both verb and gesture were less expected, both the gesture and the word were difficult to integrate to the previous sentence context. Note that in the conditions in which either the verb or the gesture were less expected, the critical verb and the overlapping gesture locally mismatched (i.e. speech, roll; gesture, walk, and vice-versa), while in the language and gesture less expected condition they locally matched (i.e. both walk). This extra manipulation allowed us to investigate and compare the effects of local and global integration of speech and gesture in sentence context.

If the immediacy assumption also applies to the unification of linguistic and extra-linguistic visual information, we expect a similar latency and amplitude

Table 1. An example of the stimulus materials. (In brackets is a verbal description of the iconic gesture. Gestures were time locked to the onset of the critical verb (underlined). ERPs were time locked to the beginning of the critical word and the gesture in each sentence. The condition coding (G+L+; G+L-, etc.) refers to the semantic fit of either the verb (language: L) or the gesture (gesture: G) to the preceding sentence context, with a minus sign indicating a semantically less expected continuation (mismatch). Less expected continuations of the preceding context are indicated in bold. Conditions B and C also contain local mismatches where the concurrent speech and gesture are different. All stimuli were in Dutch.)

- 
- (A) Language and gesture match (correct condition): L+G+  
He slips on the roof and rolls down  
(roll down)
- (B) Language mismatch: G+L-  
He slips on the roof and walks to the other side  
(roll down)
- (C) Gesture mismatch: G-L+  
He slips on the roof and rolls down  
(walk across)
- (D) Double mismatch: G-L-  
He slips on the roof and walks to the other side  
(walk across)
- 

of the N400 effect for all types of semantic 'mismatches' (i.e. language, gesture and double), revealing that the brain integrates information from both speech and gesture at the same time. Furthermore, according to the immediacy assumption, we do not expect differences across conditions with local mismatches (language and gesture mismatches) and the condition with the local match (double mismatch), since integration takes place immediately in relation to a discourse model and not in multiple steps from lower to higher levels of semantic organization. According to this view, the gesture and the concurrent speech segment (i.e. the verb) are integrated in parallel into the preceding context and not after they first formed a common semantic object.

Figure 5 shows the ERP results. In terms of their latency and amplitude characteristics, the effects are similar to the well-known N400 effect that is observed if word meaning violates the semantic context (Kutas & Hillyard 1980). However, the waveforms show a clearly biphasic morphology and the effects have a more anterior distribution than is reported for the classical N400 effect. The first negative peak in the biphasic negativity is reminiscent of the N300 that has been reported before for visual materials, and which has been found to be more negative for unrelated than for related pictures (Barrett & Rugg 1990; Holcomb & McPherson 1994; McPherson & Holcomb 1999). The N300 effect might be related to the presence of the visual-gestural information.

For the N400, an anterior distribution has been observed before for visual information such as pictures (e.g. Federmeier & Kutas 2001; West & Holcomb 2002). In the current study, the visual characteristics of the gestures might have elicited a frontal distribution. The finding that all conditions with a semantically unexpected continuation have similar topographic

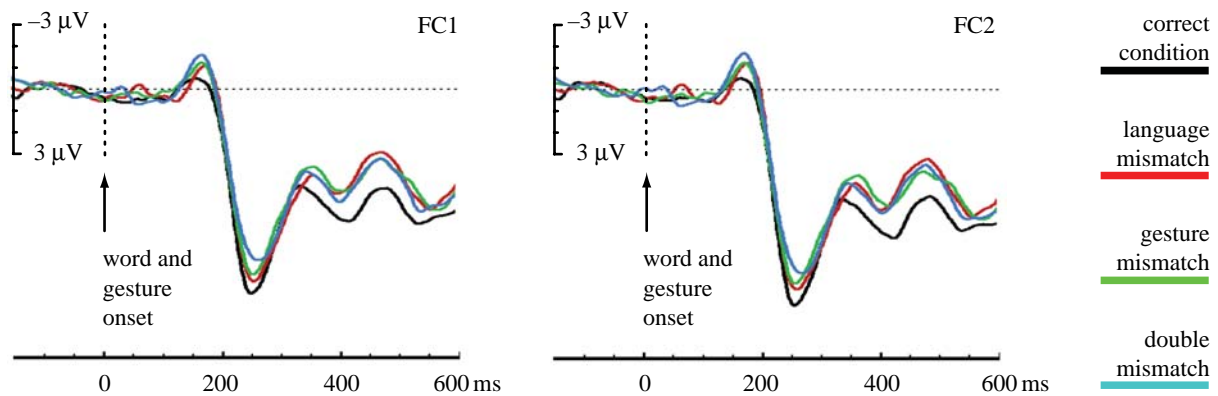


Figure 5. Grand-average waveforms for ERPs elicited in the three semantic mismatch conditions and the correct condition at two representative electrode sites (FC1 and FC2). Negativity is plotted upwards. Waveforms are time locked to the onset of spoken verb and gesture (0 ms).

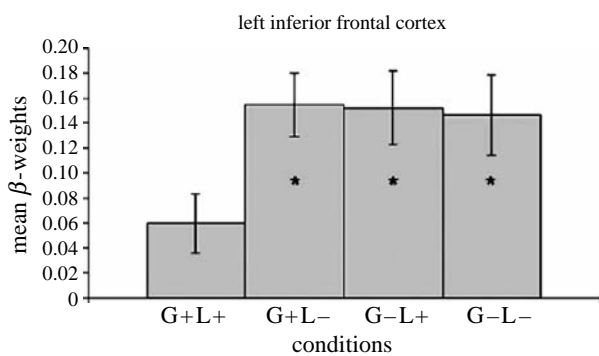


Figure 6. Gesture and speech in a sentence context. Mean activation levels ( $\beta$  weights) for the four experimental conditions in left inferior frontal cortex (BA 45/47). The activation levels are averaged over participants. An asterisk indicates a significant difference of the activation level of that condition compared with the correct condition (G+L+), at an  $\alpha$  level of  $p < 0.05$ . Error bars are standard error of the mean. G+L+, correct condition; G+L-, language mismatch; G-L+, gesture mismatch; G-L-, double mismatch.

distributions suggests that semantic integration of information from both modalities (i.e. speech and gesture) might be instantiated by overlapping neuronal sources. Interestingly, it suggests that with respect to contextual integration there is no reason to distinguish between visual semantics and verbal semantics.

Next to our ERP study, we also performed an fMRI study (Willems *et al.* *in press*), using the same stimuli in a design with the same conditions. The fMRI data (figure 6) revealed that all conditions with a semantically unexpected continuation activated the LIFG, and specifically Broca's area. This area has been claimed to be crucial for the integration of semantic information into the previous context (Hagoort 2003; 2005; Hagoort *et al.* 2004).

Together with the ERP results, the fMRI data suggest that the semantic integration of both speech and gesture semantics to sentence context involves very similar processes, and that the underlying semantic representations might be amodal in nature, in spite of the differences in input modality.

In conclusion, when understanding an utterance, the brain does not restrict itself to language information

alone, but also integrates semantic information conveyed through other modalities, such as co-speech gestures. Furthermore, the neuronal sources and the time course of the integration processes seem to be similar across gesture and language semantics. Both constrain the interpretation domain simultaneously during online processing. This opens the interesting possibility that language comprehension involves the incorporation of information in a 'single unification space' (Hagoort 2003, 2005; Hagoort *et al.* 2004), coming from a broader range of cognitive domains than is presupposed in the standard two-step model of interpretation.

## 6. MAKING SENSE OF LANGUAGE: IMMEDIATE USE OF ALL RELEVANT CONSTRAINTS

In traditional linguistic theories about meaning, a distinction is often made between the context-free rule-based combination of fixed word meanings ('sentence meaning') and the contributions made by the communicative context, such as what has been said before, who is speaking, co-speech gestures or other concomitant visual information, and the listener's background knowledge about the topic of conversation. In psycholinguistics, this analysis of meaning has evolved into the standard two-step model of language interpretation, according to which listeners (and readers) first compute a local, context-independent meaning for the sentence, and only then work out what it really means given the wider communicative context and the particular speaker.

We have discussed a wide range of ERP and fMRI findings that collectively do not sit well with this two-step model. Instead, the findings consistently point to a one-step model of language interpretation. Not only core linguistic information about the phonology, syntax and semantics of single words and sentences, but also discourse information, world knowledge and non-linguistic context information immediately conspire in determining the interpretation of compound expressions. Language input seems to be mapped onto a discourse model that takes all communicative acts, including eye gaze, iconic gestures, smiles and pointing, into consideration (Clark 1996). This is in line with the immediacy assumption, which states that these



information types are brought to bear on language interpretation as soon as they become available, without giving priority, on principled grounds, to the syntax-constrained combination of lexical-semantic information (Fregean compositionality).

Our neuroimaging findings converge with and extend behavioural observations (e.g. Trueswell & Tanenhaus 2005), and they provide support for architectures of language comprehension that allow for the rapid parallel use of multiple constraints (e.g. MacDonald *et al.* 1994; Tanenhaus & Trueswell 1995; Jackendoff 2002). Our results also converge with recent linguistic observations that the notion of context-free sentence meaning is in fact highly problematic, and that linguistic meaning is always coloured by the pragmatics of the communicative situation (Clark 1996; Perry 1997; Kempson 2001) and the wider knowledge of the world (Jackendoff 2002). The meaning of so-called 'indexicals' like 'I' and 'you', e.g. inevitably depends on who is the speaker and who is the listener (e.g. Perry 1997), and the meaning of the verb phrase 'finished X' differs, based on our world knowledge, for 'Mary finished the book' and 'the goat finished the book' (Jackendoff 2002).

Formal semantic models have been proposed that are in line with our findings. For instance, the event calculus of Van Lambalgen & Hamm (2004) assumes that the ability to construct a discourse model is derived from our ability to compute plans for achieving a given goal (Baggio *et al.* in press). This model specifies the event structure of narratives. It accounts for the fact that many core aspects of language, such as tense and aspect, really play their role beyond the sentence given at the discourse level. Moreover, one can show that even tense and aspect cannot by themselves completely determine the event structure and must recruit world knowledge (for examples, see Baggio *et al.* in press).

All this does not imply that syntax disappears in the face of discourse. Clearly, whether a language has SVO (subject verb object) or SOV (subject object verb) as its basic structure is a matter of syntax and not of semantics. Likewise, the fact that German has case morphology and English does not cannot reduce to the semantics of discourse. All we are saying here is that language processing is operating under unification principles in which linguistic information (phonology, syntax, semantics) as well as pragmatic information coming from knowledge about the context, the speaker and states of affairs in the world are handled in parallel, with a direct mapping onto an event structure (or discourse model) that goes beyond the sentence given.

Our neuroimaging studies suggest that the left inferior frontal cortex, including Broca's area, is an important node in the semantic unification network. Moreover, this area is not language specific but acts as a single unification space (as postulated in the MUC framework; Hagoort 2003, 2005; Hagoort *et al.* 2004), integrating the semantic consequences of a broader range of cognitive domains than is usually thought. Of course, the fact that various constraints on interpretation all recruit LIFG does not mean that conceptual processing during language comprehension only recruits LIFG. In fact, some recent work suggests that the resolution of referential ambiguity recruits a

very different network of brain areas (Nieuwland *et al.* in press). Crucially, however, the data reviewed here do not support the idea that some types of constraints (lexical-semantic) are handled by an early sentence-internal sense-making process whereas others (pragmatic constraints) can only be brought to bear during later computations. Knowledge about the context, concomitant information from other modalities and the speaker are immediately brought to bear on utterance interpretation, by the same fast-acting brain system that combines the meanings of individual words into a larger whole.

## 7. ENDNOTE

<sup>1</sup>Note that problems with establishing reference in discourse (i.e., finding out to what or whom a linguistic expression refers) recruit different neuronal ensembles than the two problems with meaning discussed here. Whereas lexical-semantic and world knowledge violations both generate the N400 effect and both activate LIFG, referential ambiguity elicits a sustained frontal negativity in ERPs (Nref effect; see Van Berkum *et al.* in press) and recruits a non-overlapping set of brain areas. Thus, although as suggested by the recurring N400 effects and LIFG activations, constraints from various types of domains all rapidly affect interpretation, other neural systems can also be involved in making sense of language.

## REFERENCES

- Altmann, G. T. M. 1997 *The ascent of Babel: an exploration of language, mind, and understanding*. Oxford, UK: Oxford University Press.
- Altmann, G. T. M. & Kamide, Y. 1999 Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. (doi:10.1016/S0010-0277(99)00059-1)
- Baggio, G., Van Lambalgen, M. & Hagoort, P. In press. Language, linguistics and cognition. In *Handbook of philosophy of linguistics* (eds M. Stokhof & J. Groenendijk).
- Barrett, S. E. & Rugg, M. D. 1990 Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212. (doi:10.1016/0278-2626(90)90029-N)
- Brown, C. & Hagoort, P. 1993 The processing nature of the N400: evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44.
- Caramazza, A. & Shelton, J. R. 1998 Domain-specific knowledge systems in the brain the animate–inanimate distinction. *J. Cogn. Neurosci.* 10, 1–34. (doi:10.1162/089892998563752)
- Chomsky, N. 1957 *Syntactic structures*. The Hague, The Netherlands: Mouton & Co.
- Clark, H. H. 1996 *Using language*. Cambridge, UK: Cambridge University Press.
- Crain, S. & Steedman, M. 1985 On not being led up the garden path: the use of context by the psychological parser. In *Natural language parsing* (eds D. R. Dowty, L. Karttunen & A. M. N. Zwicky), pp. 320–358. Cambridge, UK: Cambridge University Press.
- Culicover, P. W. & Jackendoff, R. 2006 The simpler syntax hypothesis. *Trends Cogn. Sci.* 10, 413–418. (doi:10.1016/j.tics.2006.07.007)
- Cutler, A. & Clifton, C. E. 1999 Comprehending spoken language: a blueprint of the listener. In *The neurocognition of language* (eds C. M. Brown & P. Hagoort), pp. 123–166. Oxford, UK: Oxford University Press.
- Ericsson, K. A. & Kintsch, W. 1995 Long-term working memory. *Psychol. Rev.* 102, 211–245. (doi:10.1037/0033-295X.102.2.211)

- Federmeier, K. D. & Kutas, M. 2001 Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *J. Exp. Psychol. Learn. Mem. Cogn.* **27**, 202–224. (doi:10.1037/0278-7393.27.1.202)
- Fodor, J. D. 1983 *The modularity of mind*. Cambridge, MA: MIT Press.
- Frege, G. 1892 Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* **100**, 25–50.
- Garrod, S. & Sanford, A. J. 1994 Resolving sentences in a discourse context: how discourse representation affects language understanding. In *Handbook of psycholinguistics* (ed. M. Gernsbacher), pp. 675–698. New York, NY: Academic Press.
- Grice, H. P. 1975 Logic and conversation. In *Syntax and semantics: speech acts* (eds P. Cole & J. L. Morgan), pp. 41–58. New York, NY: Academic Press.
- Hagoort, P. 2003 How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage* **20**, S18–S29. (doi:10.1016/j.neuroimage.2003.09.013)
- Hagoort, P. 2005 On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* **9**, 416–423.
- Hagoort, P. & Brown, C. 1994 Brain responses to lexical ambiguity resolution and parsing. In *Perspectives on sentence processing* (eds C. Clifton, L. Frazier & K. Rayner), pp. 45–81. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. 2004 Integration of word meaning and world knowledge in language comprehension. *Science* **304**, 438–441. (doi:10.1126/science.1095455)
- Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. 2003 The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* **49**, 43–61. (doi:10.1016/S0749-596X(03)00022-6)
- Hess, D. J., Foss, D. J. & Carroll, P. 1995 Effects of global and local context on lexical processing during language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **124**, 62–82.
- Holcomb, P. J. & McPherson, W. B. 1994 Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn.* **24**, 259–276. (doi:10.1006/brcg.1994.1014)
- Jackendoff, R. 2002 *Foundations of language: brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
- Jackendoff, R. 2003 Précis of foundations of language: brain, meaning, grammar, evolution. *Behav. Brain Sci.* **26**, 651–707.
- Just, M. A. & Carpenter, P. A. 1980 A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* **87**, 329–354. (doi:10.1037/0033-295X.87.4.329)
- Kempson, R. 2001 Pragmatics: language and communication. In *The handbook of linguistics* (eds M. Aronoff & J. Rees-Miller), pp. 394–427. Cambridge, MA: MIT Press.
- Kintsch, W. 1998 *Comprehension: a paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kutas, M. & Hillyard, S. A. 1980 Reading senseless sentences: brain potentials reflect semantic anomaly. *Science* **207**, 203–205. (doi:10.1126/science.7350657)
- Kutas, M. & Van Petten, C. K. 1994 Psycholinguistics electrified: event-related brain potential investigations. In *Handbook of psycholinguistics* (ed. M. A. Gernsbacher), pp. 83–143. San Diego, CA: Academic Press.
- Kutas, M., Van Petten, C. & Kluender, R. 2006 Psycholinguistics electrified II: 1994–2005. In *Handbook of psycholinguistics* (eds M. Traxler & M. A. Gernsbacher), 2nd edn. New York, NY: Elsevier.
- Lattner, S. & Friederici, A. D. 2003 Talker's voice and gender stereotype in human auditory sentence processing—evidence from event-related brain potentials. *Neurosci. Lett.* **339**, 191–194. (doi:10.1016/S0304-3940(03)00027-2)
- MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. 1994 Lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* **101**, 676–703. (doi:10.1037/0033-295X.101.4.676)
- McNeill, D. 1992 *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McPherson, W. B. & Holcomb, P. J. 1999 An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* **36**, 53–65. (doi:10.1017/S0048577299971196)
- Nieuwland, M. S. & Van Berkum, J. J. A. 2006 When peanuts fall in love: N400 evidence for the power of discourse. *J. Cogn. Neurosci.* **18**, 1098–1111. (doi:10.1162/jocn.2006.18.7.1098)
- Nieuwland, M. S., Petersson, K. M. & Van Berkum, J. J. A. In press. On sense and reference: examining the functional neuroanatomy of referential processing. *NeuroImage*.
- Osterhout, L. & Holcomb, P. J. 1992 Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* **31**, 785–806. (doi:10.1016/0749-596X(92)90039-Z)
- Osterhout, L. & Holcomb, P. J. 1995 Event-related potentials and language comprehension. In *Electrophysiology of mind* (eds M. D. Rugg & M. G. H. Coles), pp. 171–215. Oxford, UK: Oxford University Press.
- Osterhout, L., Kim, A. & Kuperberg, G. In press. The neurobiology of sentence comprehension. In *The Cambridge handbook of psycholinguistics* (eds M. Spivey, M. Joanisse & K. McRae). Cambridge, UK: Cambridge University Press.
- Otten, M. & Van Berkum, J. J. A. In press. What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Res.*
- Özyürek, A., Willems, R. M., Kita, S. & Hagoort, P. In press. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.*
- Perry, J. 1997 Indexicals and demonstratives. In *Companion to the philosophy of language* (eds C. Wright & R. Hale), pp. 586–612. Oxford, UK: Blackwell.
- Seuren, P. A. M. 1998 *Western linguistics: an historical introduction*. Oxford, UK: Blackwell.
- Sperber, D. & Wilson, D. 1986 *Relevance*. Cambridge, MA: Harvard University Press.
- Sperber, D. & Wilson, D. 1995 *Relevance: communication and cognition*. Oxford, UK: Blackwell.
- Stalnaker, R. C. 1978 Assertion. In *Syntax and semantics 9: pragmatics* (ed. P. Cole), pp. 315–332. New York, NY: Academic Press.
- St George, M., Mannes, S. & Hoffman, J. E. 1994 Global semantic expectancy and language comprehension. *J. Cogn. Neurosci.* **6**, 70–83.
- Tanenhaus, M. K. & Trueswell, C. 1995 Sentence comprehension. In *Speech, language, and communication* (eds J. L. Miller & P. D. Eimas), pp. 217–262. San Diego, CA: Academic Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. 1995 Integration of visual and linguistic information in spoken language comprehension. *Science* **268**, 1632–1634. (doi:10.1126/science.7777863)
- Trueswell, J. & Tanenhaus, M. (eds) 2005 *Approaches to studying world-situated language use: bridging the language-as-product and language-action traditions*. Cambridge, MA: MIT Press.

- Van Berkum, J. J., Hagoort, P. & Brown, C. M. 1999 Semantic integration in sentences and discourse: evidence from the N400. *J. Cogn. Neurosci.* **11**, 657–671. (doi:10.1162/089892999563724)
- Van Berkum, J. J., Zwitserlood, P., Hagoort, P. & Brown, C. M. 2003 When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cogn. Brain Res.* **17**, 701–718. (doi:10.1016/S0926-6410(03)00196-4)
- Van Berkum, J. J. A. 2004 Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? In *The on-line study of sentence comprehension: eyetracking, ERPs and beyond* (eds M. Carreiras & C. Clifton), pp. 229–270. New York, NY: Psychology Press.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V. & Hagoort, P. 2005 Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn.* **31**, 443–467. (doi:10.1037/0278-7393.31.3.443)
- Van Berkum, J. J. A., Koornneef, A. W., Otten, M. & Nieuwland, M. S. In press. Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.*
- Van Berkum, J. J. A., Van den Brink, D., Tesink, C., Kos, M. & Hagoort, P. Submitted. The neural integration of speaker and message.
- Van Lambalgen, M. & Hamm, F. 2004 *The proper treatment of events*. Oxford, UK: Blackwell.
- West, W. C. & Holcomb, P. J. 2002 Event-related potentials during discourse-level semantic integration of complex pictures. *Brain Res. Cogn. Brain. Res.* **13**, 363–375. (doi:10.1016/S0926-6410(01)00129-X)
- Willems, R. M., Özyürek, A. & Hagoort, P. In press. When language meets action: the neural integration of gesture and speech. *Cereb. Cortex.*
- Zwaan, R. A. 2004 The immersed experiencer: toward an embodied theory of language comprehension. In *The psychology of learning and motivation*, vol. 44 (ed. B. H. Ross), pp. 35–62. New York, NY: Academic Press.
- Zwaan, R. A. & Radvansky, G. A. 1998 Situation models in language comprehension and memory. *Psychol. Bull.* **123**, 162–185. (doi:10.1037/0033-2909.123.2.162)