PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2008) **363**, 1055–1069
doi:10.1098/rstb.2007.2159
Published online 21 September 2007

The fractionation of spoken language understanding by measuring electrical and magnetic brain signals

Peter Hagoort*

F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands

This paper focuses on what electrical and magnetic recordings of human brain activity reveal about spoken language understanding. Based on the high temporal resolution of these recordings, a fine-grained temporal profile of different aspects of spoken language comprehension can be obtained. Crucial aspects of speech comprehension are lexical access, selection and semantic integration. Results show that for words spoken in context, there is no 'magic moment' when lexical selection ends and semantic integration begins. Irrespective of whether words have early or late recognition points, semantic integration processing is initiated before words can be identified on the basis of the acoustic information alone. Moreover, for one particular event-related brain potential (ERP) component (the N400), equivalent impact of sentence- and discourse-semantic contexts is observed. This indicates that in comprehension, a spoken word is immediately evaluated relative to the widest interpretive domain available. In addition, this happens very quickly. Findings are discussed that show that often an unfolding word can be mapped onto discourse-level representations well before the end of the word. Overall, the time course of the ERP effects is compatible with the view that the different information types (lexical, syntactic, phonological, pragmatic) are processed in parallel and influence the interpretation process incrementally, that is as soon as the relevant pieces of information are available. This is referred to as the immediacy principle.

Keywords: speech; event-related brain potential; magnetoencephalography; N200; N400; P600/SPS

1. INTRODUCTION

Speed is one of the most remarkable characteristics of the human capacity for understanding spoken language. As listeners, we easily process three or four words per second. Although we do this seemingly without any effort, in fact a complex cascade of processes underlies our capacity for understanding. When we hear speech, numerous brain areas work together to analyse the acoustic information, select the proper words by mapping the sensory input onto stored lexical knowledge, extract the meaning of those words and integrate them into an ongoing sentential or discourse context (Marslen-Wilson 1973; Marslen-Wilson & Welsh 1978; Marslen-Wilson & Tyler 1980; Zwitserlood 1989; Norris 1994). All of these happen in a time span of only hundreds of milliseconds. An account of the neurobiology of spoken language processing can, therefore, only be adequate if the temporal dynamics is taken into consideration. In recent years, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies have given us important new insight into the network of brain areas involved in language processing. Diffusion tensor imaging has provided new data about the connectivity of language-related brain areas (Catani *et al*. 2004). However, none of these techniques has a temporal resolution of the order of milliseconds, which is necessary to study the time course of language processing. For a characterization of the temporal dynamics of spoken language comprehension, measuring electrical and magnetic brain responses is thus more appropriate. Here, I will focus on what has been learned from event-related brain potential (ERP) and magnetoencephalography (MEG) studies about the temporal and functional fractionation of the neurocognitive architecture for listening to language.

ERPs reflect the sum of simultaneous postsynaptic activity of a large population of mostly pyramidal neurons recorded at the scalp as small voltage fluctuations in the electroencephalography (EEG) time locked to sensory, motor or cognitive processes. In a particular patch of cortex, excitatory input to the apical dendrites of pyramidal neurons will result in a net negativity in the region of the apical dendrite and a positivity in the area of the cell body. This creates a tiny dipole for each pyramidal neuron, which will summate with other dipoles, provided that there is simultaneous input to the apical dendrites of many neurons and a similar orientation of these cells. The cortical pyramidal neurons are all aligned perpendicular to the surface of the cortex, and thus share their orientation. The summation of the many individual dipoles in a patch of cortex is equivalent to a single dipole calculated by averaging the orientations of the individual dipoles (Luck 2005). This equivalent current dipole is the neuronal generator (or source) of the ERP recorded at the scalp. In many cases, a

*peter.hagoort@fcdonders.ru.nl

One contribution of 13 to a Theme Issue 'The perception of speech: from sound to meaning'.

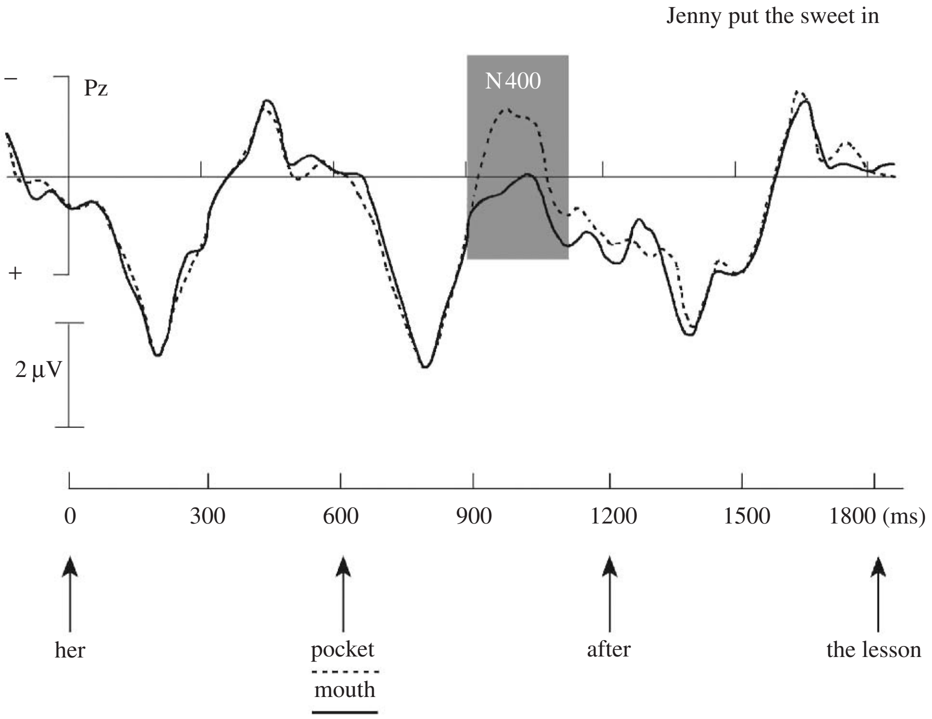

Figure 1. Modulation of the N400 amplitude as a result of a manipulation of the semantic fit between a lexical item and its sentence context. The grand average waveform is shown for electrode site Pz (parietal midline), for the best fitting word (high cloze; solid line), and a word that is less expected in the given sentence context (low cloze; dashed line). The sentences were visually presented word by word for every 600 ms. In the figure, the critical words (CWs) are preceded and followed by one word. The CW is presented at 600 ms on the time axis. Negativity is up. (Reprinted with permission from Hagoort & Brown (1994). Copyright Erlbaum.)

particular ERP component has more than one generator and contains the contribution of multiple sources. Mainly due to the high resistance of the skull, ERPs tend to spread laterally, blurring the voltage distribution at the scalp. An ERP generated locally in one part of the brain will therefore not only be recorded nearby but also at quite distant parts of the scalp. Recordings of magnetic fields instead of electrical potentials do not suffer from blurring as a consequence of high skull resistance. Since an electrical dipole is always surrounded by a magnetic field, and since these fields summate as well, a system for measuring the magnetic fields of the brain records the same activity from pyramidal neurons as EEG does. However, the spatial resolution of MEG is much greater than that of EEG, since the skull is transparent to magnetism and hence no blurring by the skull takes place. Another important difference is that MEG, in contrast to EEG, is not sensitive to dipoles with a radial orientation, i.e. dipoles are perpendicular to the skull. This holds mainly for the dipoles in the crowns of the gyri. The pyramidal cells in the banks of the sulci are oriented tangentially to the skull and their activity creates a recordable magnetic field. Thus, EEG and MEG measure overlapping but not identical overall contributions from pyramidal neurons in the cortical sheet.

In the remainder, I will focus mainly on what electrical and magnetic recordings reveal about spoken language understanding based on their high temporal resolution. In this context, I will not be able to discuss all the work that has been done, especially those on the mismatch negativity since these have mainly studied sublexical processing (Näätänen *et al.* 1997; Näätänen 2001; Pulvermüller *et al.* 2004). I will here focus on lexical processing and beyond.

2. LANGUAGE-RELEVANT EVENT-RELATED BRAIN POTENTIALS

As it holds for psycholinguistics in general, ERP research on language has been done mostly with visual input. Language-relevant ERPs have been discovered almost without exception with visual instead of spoken language input. By way of introduction, I will thus have to refer mainly to ERP studies on reading. Once the basic language-related ERP effects have been introduced, I will discuss how they contributed to the fractionation of spoken language understanding.

The electrophysiology of language as a domain of study started with the discovery by Kutas & Hillyard (1980) of an ERP component that seemed especially sensitive to semantic manipulations. Kutas & Hillyard (1980) observed a 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 1980, much has been learned about the processing nature of the N400 (for extensive overviews, see Kutas & Van Petten (1994) and Osterhout & Holcomb (1995)). As Hagoort & Brown (1994) and many others have observed, the N400 effect does not depend on a semantic violation. 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 modulate the N400 amplitude (figure 1; Hagoort & Brown 1994).

The amplitude of the N400 is most sensitive to the semantic relations between individual words, or between words and their sentence and discourse context. The better the semantic fit between a word and its context, the more reduced the amplitude of the N400. Modulations of the N400 amplitude are generally viewed as directly or indirectly related to the processing costs of integrating the meaning of a word into the overall meaning representation that is built up on the basis of the preceding language input (Osterhout & Holcomb 1992; Brown & Hagoort 1993). This holds equally when the preceding language input consists of a single word, a sentence or a discourse, indicating that semantic integration might be similar in word, sentence or discourse context (Van Berkum *et al.* 1999*b*). In addition, recent evidence indicates that sentence verification against world knowledge in long-term memory modulates the N400 in the same way (Hagoort *et al.* 2004).

In recent years, a number of ERP studies have been devoted to establishing ERP effects that can be related to the processing of syntactic information. These studies have found ERP effects to syntactic processing that are qualitatively different from the N400. Even though the generators of these effects are not yet well determined and not necessarily language specific (Osterhout & Hagoort 1999), the existence of qualitatively distinct ERP effects to semantic and syntactic processing indicates that the brain honours the distinction between semantic and syntactic processing operations. Thus, the finding of qualitatively distinct ERP effects for semantic and syntactic operations supports the claim that these two levels of language processing are domain specific. However, domain specificity should not be confused with modularity (Fodor 1983). The modularity thesis makes the much stronger claim that domain-specific levels of processing operate autonomously without interaction (informational encapsulation). Although domain specificity is widely assumed in models of language processing, there is much less agreement about the organization of the crosstalk between the different levels of processing (Boland & Cutler 1996).

ERP studies on syntactic processing have reported a number of ERP effects related to syntax (for an overview, see Hagoort *et al.* (1999)). The two most salient syntax-related effects are an anterior negativity, also referred to as LAN (left anterior negativity), and a more posterior positivity, here referred to as P600/SPS.

(a) *LAN*

A number of studies have reported negativities that are different from the N400, in that their voltage distribution over the scalp usually shows a more frontal maximum (but see Münte *et al.* 1997), sometimes larger over the left than the right hemisphere, although in many cases the distribution is bilateral (Hagoort *et al.* 2003). Moreover, the conditions that elicit these frontal negative shifts seem to be more strongly related to syntactic processing than to semantic integration. Usually, LAN effects occur within the same latency range as the N400, i.e. between 300 and 500 ms post stimulus (Osterhout & Holcomb 1992; Kluender & Kutas 1993; Münte *et al.* 1993; Rösler *et al.* 1993; Friederici *et al.* 1996). However, in some cases, and almost exclusively with spoken language input, the latency of a left frontal negative effect is reported to be much earlier, somewhere between approximately 100 and 300 ms (Neville *et al.* 1991; Friederici *et al.* 1993; Friederici 2002).

In some studies, LAN effects have been reported for violations of word-category constraints (Münte *et al.* 1993; Friederici *et al.* 1996; Hagoort *et al.* 2003). That is, if a word of a different syntactic class is presented (e.g. a verb) instead of the required one (e.g. a noun in the context of a preceding article and adjective), early negativities are observed. Friederici and colleagues (Friederici 1995; Friederici *et al.* 1996) have tied the early negativities specifically to the processing of word-category information. However, sometimes similar early negativities are observed with number, case, gender and tense mismatches in morphologically rich languages (Münte *et al.* 1993; Münte & Heinze 1994).

LAN effects have also been related to verbal working memory in connection with filler-gap assignment (Kluender & Kutas 1993). This working memory account of the LAN is compatible with the finding that lexical, syntactic and referential ambiguities seem to elicit very similar frontal negativities (Hagoort & Brown 1994; King & Kutas 1995; Van Berkum *et al.* 1999*a*; Kaan & Swaab 2003). Lexical and referential ambiguities are clearly not syntactic in nature, but can be argued to tax verbal working memory more heavily than sentences in which lexical and referential ambiguities are absent. Future research should indicate whether or not these two functionally distinct classes of LAN effects can be dissociated at a more fine-grained level of electrophysiological analysis.

(b) *P600/SPS*

A second ERP effect that has been related to syntactic processing is a later positivity, nowadays referred to as P600 or P600/SPS (Osterhout *et al.* 1997; Coulson *et al.* 1998; Hagoort *et al.* 1999). One of the antecedent conditions of P600/SPS effects is a violation of a syntactic constraint. If, for instance, the syntactic requirement of number agreement between the grammatical subject of a sentence and its finite verb is violated (see (1) below, with the critical verb form in italics; the * indicates the ungrammaticality of the sentence), a positive-going shift is elicited by the word that renders the sentence ungrammatical (Hagoort *et al.* 1993). This positive shift starts at approximately 500 ms after the onset of the violation and usually lasts for at least 500 ms. Given the polarity and the latency of its maximal amplitude, this effect was originally referred to as the P600 (Osterhout & Holcomb 1993) or, on the basis of its functional characteristics, as the syntactic positive shift (SPS; Hagoort *et al.* 1993).

(1) *The spoilt child *throw* the toy on the ground.

An argument for the independence of this effect from possibly confounding semantic factors is that it also occurs in sentences where the usual semantic/pragmatic constraints have been removed (Hagoort & Brown 1994). This results in sentences like (2a) and (2b) where one is semantically odd but grammatically correct, whereas the other contains the same agreement violation as in (1):

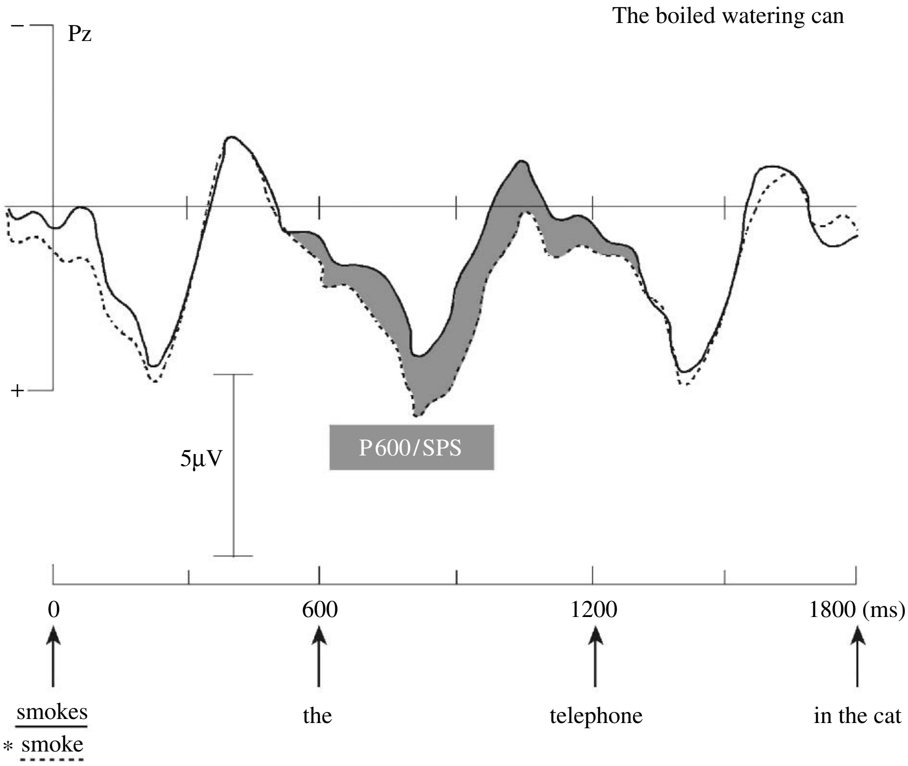

Figure 2. ERPs to visually presented syntactic prose sentences. These are sentences without a coherent semantic interpretation. A P600/SPS is elicited by a violation of the required number agreement between the subject noun phrase and the finite verb of the sentence. The averaged waveforms for the grammatically correct (solid line) and the grammatically incorrect (dashed line) words are shown for electrode site Pz (parietal midline). The word that renders the sentence ungrammatical is presented at 0 ms on the time axis. The waveforms show the ERPs to this and the following two words. Words were presented word by word, with an interval (stimulus onset asynchrony) of 600 ms. Negativity is plotted upwards; asterisk indicates the syntactic isolation. (Reprinted with permission from Hagoort & Brown (1994). Copyright Erlbaum.)

(2a) The boiled watering can *smokes* the telephone in the cat.
(2b) * The boiled watering can *smoke* the telephone in the cat.

If one compares the ERPs with the italicized verbs in (2a) and (2b), a P600/SPS effect to the ungrammatical verb form is visible (figure 2). Despite the fact that these sentences do not convey any conventional meaning, the ERP effect of the violation demonstrates that the language system is nevertheless able to parse the sentence into its constituent parts.

Similar P600/SPS effects have been reported for a broad range of syntactic violations in different languages (e.g. English, Dutch, German), including phrase-structure violations (Neville *et al.* 1991; Osterhout & Holcomb 1992; Hagoort *et al.* 1993), subcategorization violations (Osterhout *et al.* 1994, 1997; Ainsworth-Darnell *et al.* 1998), violations in the agreement of number, gender and case (Hagoort *et al.* 1993; Osterhout & Mobley 1995; Münte *et al.* 1997; Osterhout 1997; Coulson *et al.* 1998), violations of subjacency (Neville *et al.* 1991; McKinnon & Osterhout 1996) and of the empty-category principle (McKinnon & Osterhout 1996).

Recently, a P600/SPS is also reported in relation to thematic role assignment (Kuperberg *et al.* 2003; Van Herten *et al.* 2004; Kim & Osterhout 2005; Wassenaar & Hagoort 2007). In this case, a P600/SPS is elicited when constraints for grammatical role assignment are in conflict with thematic role biases. For instance, Kim & Osterhout (2005) report a P600/SPS to the verb *devouring* in the sentence 'The hearty meal was devouring …', where the first noun phrase (NP) is not a good *agent*, but would be fine as a *theme*. The fascinating possibility suggested by these results is that a strong thematic bias could induce a tendency to detect a grammatical error where there is none ('-ing should be -ed') or to assign the grammatical role of object to the first NP, whereas the syntactic cues indicate that it is the subject of the sentence. These opposing tendencies result in a P600/SPS.

In summary, two classes of syntax-related ERP effects have been consistently reported. These two classes differ in their polarity, topographic distribution and latency characteristics. In terms of latency, the first class of effects is an anterior negativity. Apart from LANs related to working memory, anterior negativities to syntactic violations are mainly seen. In a later latency range, positive shifts occur that are not only elicited by syntactic violations but also in grammatically well-formed sentences that vary in complexity (Kaan *et al.* 2000), as a function of the number of alternative syntactic structures that are compatible with the input at a particular position in the sentence (syntactic ambiguity; Osterhout *et al.* 1994; Van Berkum *et al.* 1999*a*), or when constraints for grammatical role assignment are overwritten by thematic role biases (Kim & Osterhout 2005).

3. SPEECH-RELATED EVENT-RELATED BRAIN POTENTIAL EFFECTS

Despite the fact that spoken language is the primary mode of language communication, and reading a recent invention that is usually only acquired with substantial effort and through formal instruction, the majority of psycholinguistic research is on reading rather than listening or speaking. Likewise, most neuroimaging and ERP/MEG studies on language have used visual instead of auditory input. Hence, the number of ERP studies on spoken language processing is relatively limited. Studies that present natural connected speech have often found the same basic semantic (N400) and syntactic (P600/SPS) ERP effects as in reading, although the overall morphology of the ERP waveforms is quite different in speech than in reading.

(a) *Auditory N400/N200*

Regarding the N400 effect in speech, a number of studies have found the effect to have an earlier onset latency than in reading and usually the effect has a longer duration for speech than for written input (Holcomb & Neville 1990, 1991; Hagoort & Brown 2000*a*). For instance, Holcomb & Neville (1990) reported a very early effect to semantic anomalies in sentences spoken at a normal rate. They characterized this as an N400 effect. Over occipital sites, the onset latency of the auditory N400 effect was as early as 50–100 ms, and 150 ms over left and right parietal sites. The earlier onset of N400 effects on connected speech is surprising since, in contrast to written words, spoken words are encoded in a signal that is extended over time.

However, there is some doubt as to whether the auditory N400 effect is indeed a similar monophasic negative shift as the visual N400 effect is. The alternative is that the so-called auditory N400 effect actually is composed of at least two separate negative polarity effects, of which only the second negativity is an N400. In one of the first ERP studies on spoken language processing, McCallum *et al.* (1984) already reported that in their data the auditory N400 effect seemed to be preceded by a separate N200 effect. This early effect reached its maximal amplitude for the semantically incongruous sentence endings between 208 and 216 ms. The possible separation of the overall negative shift into a functionally different early and late negative effect was tested in a series of studies by Connolly *et al.* (1990, 1992, 1995). These authors compared ERPs with sentence-final words in highly constraining sentence contexts (e.g. 'The king wore a golden crown.') with ERPs in sentence contexts with low constraints (e.g. 'The woman talked about the frogs.'). A negative shift to words of low constraining sentences was observed (i.e. frogs) relative to words in highly constraining sentences (i.e. crown). The authors reported (Connolly *et al.* 1990) that individual difference waveforms (but not the grand averages), obtained by subtracting for each subject from each other the waveforms to the critical words (CWs) in the two context conditions, showed two distinct peaks, an early one with a central distribution (N200 effect) and a later one with a centroparietal distribution (N400 effect). The authors suggest that the N200 effects reflect acoustic/phonological word processing. That is, if the initial phoneme of the CW mismatches with the onset of the expected word, an N200 effect emerges. The N400 amplitude is claimed to be modulated by semantic expectancy.

Connolly & Phillips (1994) have attempted a more direct test of their account of an early and a later negativity in the ERP for words that do not allow a straightforward semantic fit with the context. In their study, next to semantically correct sentences, they presented sentences that ended in a semantically anomalous way. The anomalous word either started with the same phoneme as the most expected word given the sentence context (phoneme match condition), or its onset was different from the expected ending (phoneme mismatch condition). One negative peak (the N400) was reported for the phoneme match condition, whereas two negative peaks were observed for the phoneme mismatch condition. The authors attributed the earlier negativity to the phonemic deviation from the expected lexical form. Therefore, they called this effect the phonological mismatch negativity (PMN). The account of this early negative shift as a PMN is based on the idea that in spoken word recognition, word-initial sounds activate a cohort (Marslen-Wilson & Tyler 1980) or a shortlist (Norris 1994) of possible lexical candidates. In the process of recognizing a word, further incoming sensory information and top-down contextual information result in a reduction of the cohort or shortlist of possible candidates to one. This is the word that is actually perceived. Since in the phoneme mismatch condition the expected word is not a member of the cohort of lexical candidates, the mismatch can be detected early. In contrast, in the phoneme match condition, the expected word is a member of the cohort of lexical candidates that is instantiated by the onset of the sentence-final anomalous word. Therefore, the mismatch supposedly can be detected only later, when context information contributes to the pruning of the cohort of lexical candidates.

Hagoort & Brown (2000*a*) reported two experiments on semantic violations in connected speech that both resulted in substantial N400 effects time-locked to the word in the sentence that was semantically at odds with the preceding context. Unlike the prototypical visual N400 effect, which tends to be slightly larger over the right hemisphere, the auditory N400 effect was either symmetrical or larger over the left than the right hemisphere. However, just as in the visual modality, the auditory N400 effect had a clear posterior distribution. Both functionally and topographically, there is a strong correspondence between the visual and the auditory N400 effects. The hemispheric differences suggest that probably there is also a contribution from non-overlapping neuronal generators for the two input modalities.

In the Hagoort & Brown (2000*a*) study, the onset of N400 effects in the auditory modality was similar to that in the visual modality. However, it was found that, in addition, the anomalies elicited another effect that preceded the N400 effect in time. The onset of this effect (150 ms) was quite comparable to what Holcomb & Neville (1991) reported for their parietal

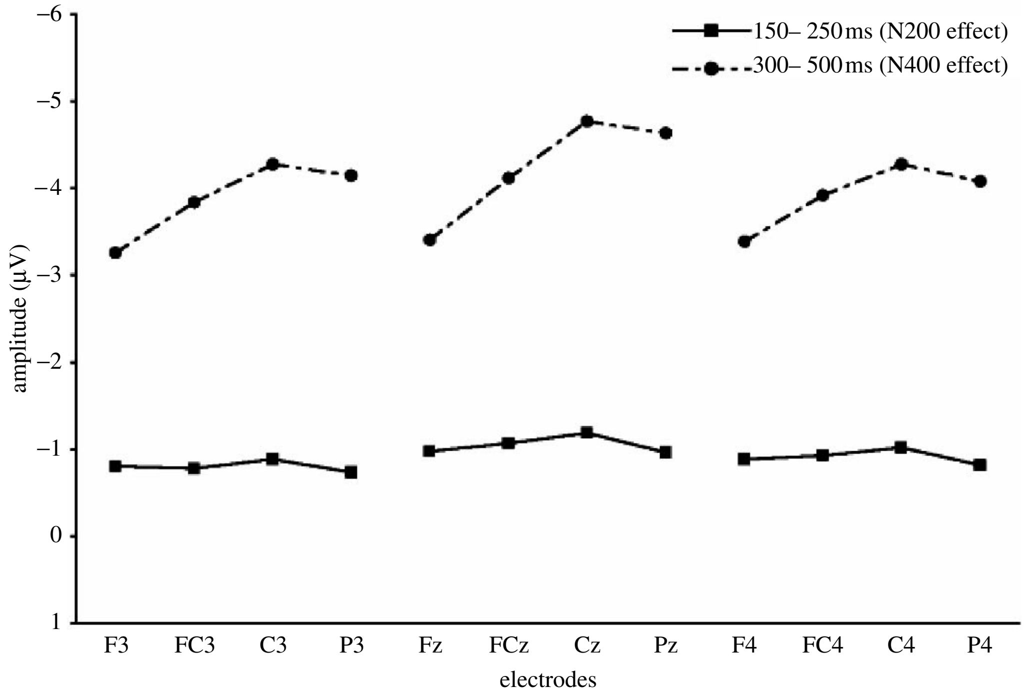

Figure 3. Connected speech. Distribution of the N200 and N400 effects for four left hemisphere sites (F3, FC3, C3 and P3), four midline sites (Fz, FCz, Cz and Pz) and four right hemisphere sites (F4, FC4, C4 and P4). The N200 effect was determined by subtracting the mean amplitude in the 150–250 ms latency window of the grand average ERP for the semantically congruent sentence-final words from the mean amplitudes of the grand average ERP for the semantically anomalous sentence-final words that did not share the same initial phonemes as the semantically congruous words. The N400 effect was determined in the same manner on the basis of the mean amplitudes in the 300–500 ms latency window (Van den Brink *et al.* 2001).

sites, and to the PMN of Connolly & Phillips (1994). Van den Brink *et al.* (2001) found further evidence for an early N200 effect and a later N400 effect. The early effect occurred in the latency range of 150–250 ms and had an even distribution over the scalp. The N400 effect between 300 and 500 ms had the usual posterior distribution (figure 3). Moreover, in contrast to the N400 effect, the N200 effect disappeared when the semantic anomaly shared the initial CV cluster with the semantically expected word.

The design of the study by Van den Brink *et al.* (2001) was very similar to that of the study by Van Petten *et al.* (1999). These authors presented sentences that ended with a pause followed by a target word that was either (i) the word with the highest cloze probability (e.g. 'It was a pleasant surprise to find that the car repair bill was only seventeen *dollars*.'), or a semantically incongruous ending that consisted of (ii) a word that shared the initial phonemes with the congruous target (e.g. *dolphins*), (iii) a word that had a different onset but rhymed with the highest cloze probability word (e.g. *scholars*), or (iv) a semantically incongruous sentence-final word without any form overlap with the congruous target (e.g. *bureau*). All three semantic anomaly conditions resulted in a significantly larger N400 than the congruous (highest cloze probability) ending. However, the onset of the N400 effect differed between conditions, as a function of the phonological overlap with the congruous ending. The onset of the N400 effect was later for the anomaly with a word-initial form overlap with the congruous ending, compared with onset latencies for the fully anomalous and rhyme word conditions. This suggests that the onset of the N400 effects was determined by the moment at which the acoustic input became inconsistent with a congruous sentence completion.

In §3*b*, how N200 and N400 effects can be characterized in terms of a functional account of spoken word recognition is discussed.

(b) *Aspects of spoken word recognition*

In models of spoken word recognition, usually a distinction is made between access, selection and integration (Marslen-Wilson 1987). The mental lexicon is the crucial interface between language form and content, two fundamentally distinct knowledge domains (Marslen-Wilson 1989). Lexical access refers to the mapping of the input signal onto word form representations in the mental lexicon. Selection refers to the discrimination among competing alternatives for a match between the acoustic input and lexical form representations (Frauenfelder & Tyler 1987). Through the mapping dynamics of access and selection, the retrieval of information associated with the word form is achieved, including the syntactic properties (e.g. gender, word class) and the meaning of a lexical item. However, since we normally perceive words not in isolation, but in the context of other words, the retrieved syntactic and semantic information have to be integrated with the higher-level context representation of the preceding part of utterance. We will refer to this process as integration (cf. Marslen-Wilson 1987). The nature of spoken word recognition is co-determined by the specifics of the speech signal. Mapping the spoken signal onto information in the mental lexicon is in many ways very different from mapping a visual signal onto lexical information. A central aspect of processing spoken words is that it occurs from left to right, starting from word onset (Marslen-Wilson & Welsh 1978). This left-to-right processing of spoken words allows the identification of the moment in time at which a particular word is recognized. The recognition point (RP) is defined as that part of the signal where the actual word becomes uniquely different from all other words in the listener's mental lexicon (Marslen-Wilson 1987). For instance, when presented in isolation, the RP of the word *captain* occurs after the /t/, since it is at this point that the sensory information excludes the only remaining alternative word candidate *captive*. For most multisyllabic words, the RP is located well before the end of the word (Marslen-Wilson 1984, 1987).

Electrophysiological evidence supporting the behavioural evidence for the concept of a RP in spoken word recognition was recently found in a study by O'Rourke & Holcomb (2002). These authors used the well-established fact that each content word elicits an N400 component. They had subjects listen to words with an early RP (e.g. pupil) and words with a late RP (e.g. carriage), without any additional task. When measured from acoustic word onset, the N400 component had an earlier peak latency for words with early RPs compared with words with late RPs. However, when the ERPs were averaged time-locked to the individual RPs, this difference disappeared, and the waveforms were identical following the RP.

In sentence context, it has been found that context information can speed up word processing (Tyler & Wessels 1983; Zwitserlood 1989). On the basis of experimental evidence, it is estimated that for selecting the word forms of one- and two-syllable content words, subjects need to hear an average of 200 ms in sentence context and more than 300 ms in isolation (Grosjean 1980; Marslen-Wilson 1984).

Although the amplitude of the N400 component is larger for content words than function words, and smaller for high frequency than low frequency words, modulations of its amplitude (i.e. the N400 effect) seem to be especially sensitive to integration—that is, to match the content specifications of a lexical candidate against the content specifications of the word, sentence or discourse context (Brown & Hagoort 1993; Holcomb 1993; Kellenbach & Michie 1996; Hagoort *et al.* 2004).

In contrast to the N400, the N200 effect seems specific for spoken language. The findings of Van den Brink *et al.* (2001) suggest a different functional account for the N200 effect than for the N400 effect. They found no N200 when the semantic anomaly shared its phonological onset with the contextually expected word candidate. Based on the initial form overlap, the expected candidate might initially be co-activated as a member of the cohort or shortlist of activated lexical candidates. This cohort/shortlist therefore contained the semantic features that are supported by the context. The authors propose that the amplitude modulation of the early negativity preceding the N400 effect reflects the lexical selection process that occurs at the interface of lexical form and contextual meaning. The word-initial speech segment activates a cohort of lexical candidates that are compatible with the initial stretch of the speech signal. This is a purely form-driven bottom-up process. Once the initial cohort (or shortlist) of lexical candidates is instantiated, top-down context information can start to have its effect. The interaction of form-based activation and content-based modulatory influences on the activational status of the available lexical candidates results in the selection of the lexical candidate that is optimally compatible with both form and content constraints. The N200 effect might reflect the lexical selection process that occurs at the interface of lexical form and contextual meaning. That is, if the contextual specifications do not support the form-based activation of a lexical candidate, an N200 effect is visible relative to a situation in which form-based activation is supported by contextual specifications. A large N200 would then indicate that the cohort of activated candidates does not contain semantic features that fit the preceding sentence frame well.

In contrast to the N200 effects, the N400 effect is claimed to arise at the content level only. Once a word's meaning is activated, the language processing system tries to incorporate its content specifications into the overall higher-order representation of the preceding utterance part. Even a clear mismatch between the meaning of a word and the semantics of the context does not prevent the mandatory process of matching the semantics of a word against the semantics of the context. It is this purely content matching and integration process of the most highly activated lexical candidates against their context that is reflected in the amplitude of the N400. The better the semantic fit, the more reduced the N400 amplitude.

One caveat has to be made with respect to these functional accounts. Not always is the N200 observed when it should be expected on the basis of the above account (Van Petten *et al.* 1999). This might be due to the overlapping component problem. The N200 and the N400 effects tend to overlap in time. This overlap implies that it is very hard to disentangle the neuronal generators of these two effects, and hence to find solid evidence, showing that indeed the N200 and N400 effects are qualitatively distinct.

(c) *Lexical selection versus integration*

The temporal resolution of EEG/MEG also allows one to investigate whether lexical selection is a prerequisite for the process of integration. This issue was addressed in a recent study by Van den Brink *et al.* (2006), which was designed to investigate the temporal relationship between lexical selection and semantic integration in auditory sentence processing. The authors investigated whether there is a discrete moment when lexical selection ends and semantic integration begins, or whether these two processes are of a cascading nature with semantic integration starting before lexical selection is completed. Information about the RP was used to investigate the onset of the N400 effect. Preceding the ERP experiment, a gating study was done on 522 spoken words. The gating method (Grosjean 1980; Tyler & Wessels 1983) allows the presentation of incremental portions of the acoustic signal until the full word is presented. For each gate, subjects specify which word they believe to be listening to and how confident they are that their response is

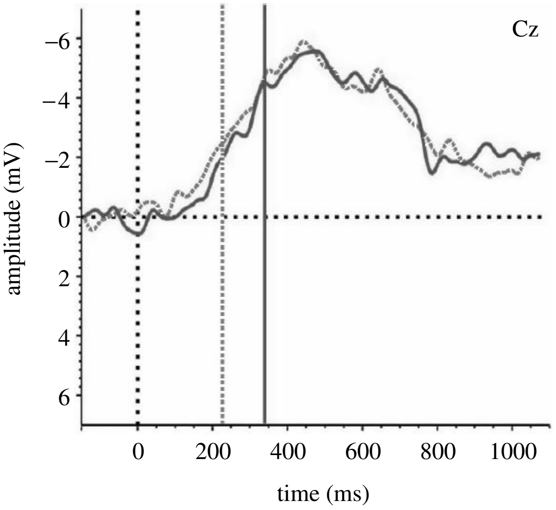

Figure 4. Connected speech. Difference waveforms from a representative electrode site (Cz), for semantically congruent versus semantically anomalous words with early RPs (dotted line; mean 229 ms) and late RPs (solid black line; mean 343 ms). The time axis is in milliseconds. Negative polarity is plotted upwards.

correct. In this way, the RP can be established empirically. This study revealed that whereas the duration of the 522 words was on average 516 ms, their mean RP was well before word offset, namely at 286 ms. These words served as congruent and incongruent completions of spoken sentences. The results of the gating study were used to divide the CWs into two groups. One group contained contextually congruent and incongruent words with early RPs (with a mean of 230 ms), and the other group consisted of congruent and incongruent completions with late RPs (mean of 340 ms). If integration is dependent on word identification, then the N400 effect should set in at or soon after the RP and should therefore differ between the early and late RP words. However, if semantic integration starts before the word is uniquely recognized on the basis of bottom-up acoustic information, then the onset of the N400 could occur prior to the RP and should not have to differ as a function of RP latency. The results revealed that, despite a mean difference in RP of at least 100 ms between words with early and late RPs, the factor of early or late RPs neither affected the onset nor the peak of the N400. Incongruent completions in both groups elicited an N400 before the RP, which had an onset latency at approximately 200 ms (figure 4). This indicates that irrespective of whether the selection process had successfully singled out one candidate, integration processing was started up nonetheless. Moreover, integration starts very early, well before the full word has been heard and in many cases even before the RP. These results favour a cascading account of spoken word processing in context. It seems that the semantic integration process does not wait until one appropriate candidate has been selected on the basis of a phonological analysis (cf. Marslen-Wilson (1989) for a similar view).

Up until recently, many researchers have shared the view that language processing is continuous and incremental, but that each incoming word is processed with bottom-up priority, and a number of spoken word processing models, such as the cohort model and shortlist, have incorporated this view (Marslen-Wilson & Welsh 1978; Marslen-Wilson 1987, 1993; Norris 1994). The findings of Van den Brink *et al.* (2006) could be explained in terms of bottom-up priority of spoken word processing. However, a number of recent ERP studies have investigated the possibility that expectancies for an upcoming word form are being generated on the basis of the preceding context in combination with the comprehender's knowledge about the world (Wicha *et al.* 2004; DeLong *et al.* 2005; Van Berkum *et al.* 2005). The results of these studies revealed that, in highly constraining contexts, words are not only rapidly integrated into the higher-order meaning representation of the preceding sentence or discourse context but also that the constraining context is used to form probabilistic predictions of which specific word will be presented next. This idea of anticipation is not new. Several models of spoken word recognition such as the Logogen model by Morton (1969) and TRACE by McClelland & Elman (1986) have already allowed for lexical preactivation of words based on the context. However, compelling evidence for preactivation during online sentence processing is limited, whereas a number of behavioural studies have provided evidence for an initial bottom-up priority based on the acoustic input (for a review of the literature relevant to this issue, see Van Berkum *et al.* (2005)).

In the case of a violation of anticipation of specific words, it is not surprising to find that the N400 effect sets in before the RP. Analysis of the first phonemes reveals that they do not match with those of any of the words anticipated. However, in light of the majority of behavioural evidence favouring bottom-up priority for acoustic processing of initial phonemes of the perceived word, Van den Brink *et al.* (2006) propose the following scenario. During sentence processing, there is a certain time frame in which lexical selection on the basis of a combination of the acoustic analysis of a word's first phonemes and context-based specifications can take place, and presumably it happens in the case of congruent words in highly to moderately constraining sentences. However, when selection of one appropriate candidate is difficult, as would be the case for anomalous words, or even congruent words in low-constraining contexts, integration as reflected by the N400 seems to be attempted for those candidates that match the acoustic input.

In conclusion, when words are spoken in context, there is no 'magic moment' when lexical selection ends and semantic integration begins. Irrespective of whether words have early or late RPs, semantic integration processing is initiated before words can be identified on the basis of the acoustic information alone.

(d) *Neuronal generators of the N400*

In an MEG version of the Van den Brink and Hagoort study (Jensen *et al.* in preparation), it was found that the N400m effect is strongly left lateralized, with maximal activity over temporo-frontal areas (see figure 5). MEG studies have identified a strong and a

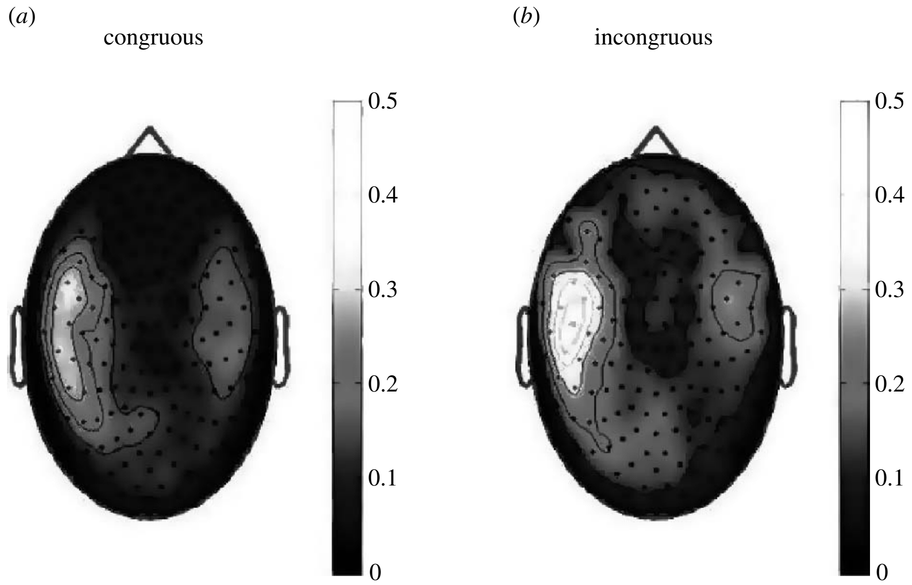

Figure 5. MEG recordings of the N400. The N400m component is observed from 300 to 600 ms. The topography of the evoked fields (planar gradients of the evoked fields) for (*a*) the congruent nouns and (*b*) the incongruent nouns in sentence context. For the incongruous nouns, an increased N400m is visible over the left hemisphere.

weak source of the N400m in the left and right superior temporal sulci, respectively (Helenius *et al.* 1998; Halgren *et al.* 2002; Helenius *et al.* 2002). Using distributed currents source modelling, Halgren *et al.* (2002) identified an additional source in the left prefrontal cortex. The sources in the superior temporal sulci are consistent with the topography of the N400m effect in figure 5. It is interesting to note that the ERP study (Van den Brink *et al.* 2006) using the same materials did not show a left lateralization of the N400 effect. Nevertheless, the topographies of the N400 effect are consistent for the EEG and MEG studies. The topography of the N400 ERP effect can be explained by dipolar sources in the left and right temporal sulci, given that these dipoles are approximately oriented towards the vertex of the head. Since both dipoles produce negative potentials over the midline, no lateralization is observed.

An MEG component with a similar time course and distribution is the M350 (Embick *et al.* 2001). The latency of the M350 has been found to be sensitive to word frequency. The M350 has been recorded mainly for written words. For the N400, it is the amplitude rather than the latency that is responsive to frequency. It is currently unclear if and, if so, to what degree the M350 and the N400m are based on overlapping neuronal generators.

4. BEYOND THE SINGLE UTTERANCE

A major task of psycholinguistics is to find out how syntactic, semantic and referential analyses are orchestrated as language comprehension unfolds in time. Research on this topic has primarily focused on the relative timing and the informational dependency of these various analyses. An important unresolved question, for example, is whether the different types of analysis are conducted in some principled sequential order, with some theorists assigning a fundamental priority to syntactic analysis (Frazier 1987; Mitchell *et al.* 1995) and others instead arguing for a simultaneous evaluation of syntactic, semantic and referential aspects of the input (Tyler & Marslen-Wilson 1977; Marslen-Wilson & Tyler 1980; MacDonald *et al.* 1994; Tanenhaus & Trueswell 1995; Jackendoff 1999; Hagoort 2005). A closely related question is whether the results of semantic and referential processing affect the initial syntactic analysis of the input (Crain & Steedman 1985; Altmann 1988) or not (Ferreira & Clifton 1986).

Van Berkum *et al.* (2003) explored the possibility that the ERP method can also be used to selectively track some of the *referential* aspects of language comprehension, while people listen to a piece of discourse. This is a relatively unexplored territory in electrophysiological research. Although ERPs have been used to address issues in referential processing before (Osterhout & Mobley 1995; Osterhout *et al.* 1997; Streb *et al.* 1999; Schmitt *et al.* 2002; see Kutas *et al.* 2000, for a brief review), no studies have directly looked for an ERP signature of referential analysis in discourse-level spoken language comprehension.

In the Van Berkum *et al.* (2003) study, subjects were asked to listen to several ministories, such as the one below (translated from Dutch, boldface and italics added):

(3) David had told *the boy and the girl* to clean up their room before lunchtime. But the boy had stayed in bed all morning, and the girl had been on the phone all the time. David told **the girl** that had been on the phone to hang up.

Following earlier research on this topic (Crain & Steedman 1985), the stories were varied such that they provided either a *single* unique referent for the NP 'the girl', as in (3), or *two* equally eligible referents, as in (4):

(4) David had told *the two girls* to clean up their room before lunchtime. But one of the girls had stayed in bed all morning, and the other had been on the

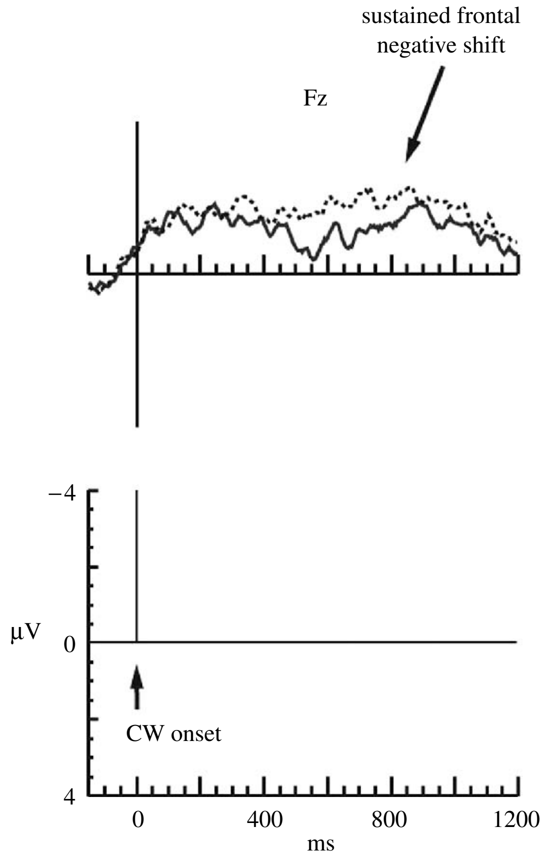

Figure 6. Grand average ERP waveform elicited by spoken singular nouns presented in a one-referent (solid line) and a two-referent (dotted line) context. The acoustic onset of the noun is at 0 ms, and negative polarity is plotted upwards. The waveforms are shown for a representative frontal midline site (Fz).

phone all the time. David told **the girl** that had been on the phone to hang up.

The authors found that referentially ambiguous spoken nouns elicited a negative shift that emerged in the ERPs at approximately 300–400 ms after acoustic onset, and that, although widely distributed, was most prominent and most sustained at anterior recording sites (figure 6). These findings show that very quickly the processing system has determined whether a singular definite noun has a single unique referent in the earlier discourse or not (in italics in (3) and (4), respectively).

The higher-level processes associated with establishing reference, particularly those that require consulting one's model of the prior discourse, are frequently assumed to be rather slow, as compared with the lower-level syntactic and sentence-semantic aspects of comprehension (Kintsch 1998). However, listeners needed at most only 300–400 ms at some level to detect a difference between, say, 'the girl' in a discourse that had introduced a single girl or that had introduced *two* girls. Given that all known sentence-semantic and sentence-syntactic ERP effects emerge within some 500 ms after CW onset (see Brown *et al.* (2000) for review), this discourse-related ERP effect occurs within the same temporal window of opportunity. Of course, this does not mean that referential ambiguity is *always* detected within some 300–400 ms. For instance, the moment at which an unfolding noun reveals its number depends on a wide variety of factors, including where the language at hand codes a noun for its number (e.g. suffix, prefix or other); the duration of the spoken noun at hand; and the way the stem of a noun changes with pluralization. In sum, the early onset of the referential ERP effect should primarily be taken as an indication that discourse-dependent referential ambiguity *can* be detected within that short duration.

Apart from its rapid emergence, the referentially induced ERP effect observed with spoken language is also 'immediate' in suggesting that reference is established *incrementally*, i.e. at each relevant word coming in. In Dutch, a referentially ambiguous NP can always be extended by a post-nominal modifier that supplies additional information, such as a relative clause ('the girl that was waiting'). In principle, the comprehension system might thus delay its attempt to establish reference until later sentential input signals that the NP is unequivocally complete. What these and earlier (Van Berkum *et al.* 1999*a*) findings suggest is that the comprehension system does *not* do this, and instead initiates sufficient referential processing at the head noun to at least determine, within some 300–400 ms after noun onset, whether it is referentially unique or not.

In addition to timing, there is also information on the *nature* of the referentially induced ERP effect, a frontally dominant and sustained negative shift. The frontally sustained negative shift reported for spoken referentially ambiguous nouns is very similar to sustained ERP effects observed under conditions of increased memory load in language comprehension (Kluender & Kutas 1993; King & Kutas 1995; Friederici *et al.* 1996; Kutas 1997; Münte *et al.* 1998; Fiebach *et al.* 2001; Vos *et al.* 2001) as well as in non-linguistic processing tasks (Rösler *et al.* 1993; Donaldson & Rugg 1999; Rugg & Allan 2000).

It is not difficult to imagine why referential ambiguity might be a memory-demanding situation. For one, referential ambiguity may trigger additional retrieval from episodic discourse memory, associated with a search for less obvious clues that might help to infer the most plausible referent (Myers & O'Brien 1998). Alternatively, referential ambiguity may require the system to actively maintain two candidate fillers for an unresolved single referential slot in working memory (see Gibson (1998) for an account of referentially induced working memory load in sentence comprehension). The latter would explain why the ERP effect of referential ambiguity resembles the ERP effect elicited by various other types of expressions that impose a higher load on working memory, such as (i) object-relative clauses (King & Kutas 1995; Kutas 1997), (ii) temporal expressions like '*Before* the psychologist submitted the article, the journal changed its policy' (Münte *et al.* 1998) in which the information supplied in the first phrase does not describe what actually happened first, or (iii) expressions like 'The *pitcher* fell down and broke/cursed' that contain a lexically ambiguous word (Hagoort & Brown 1994).

In the domain of language processing, memory-related sustained frontal negativities are sometimes referred to as left anterior negativities or LAN effects,

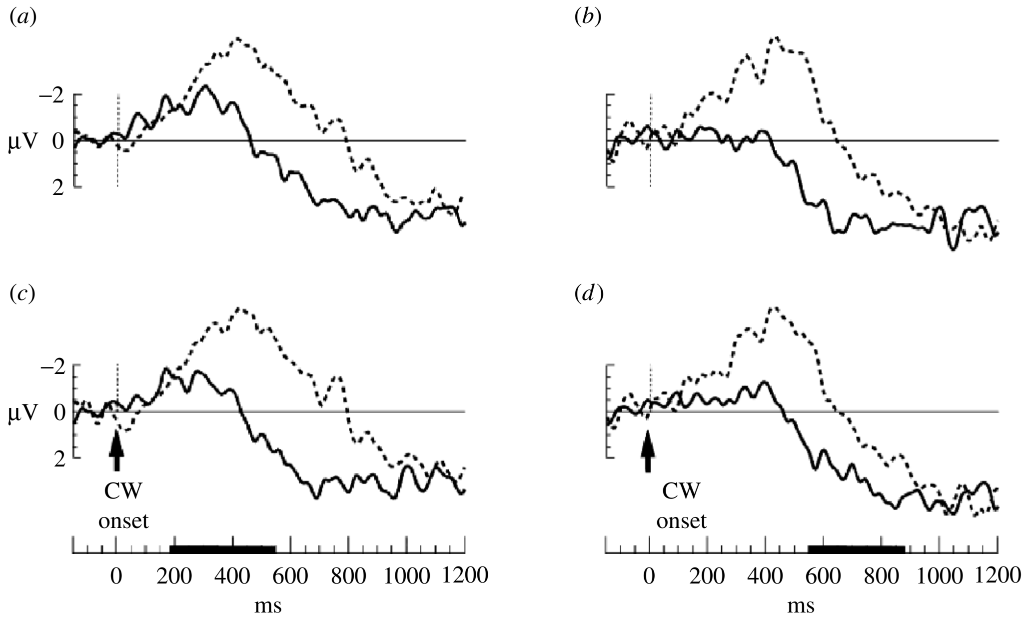

Figure 7. Discourse-dependent semantic anomaly effects by word position and length. Grand average ERPs, at Pz, elicited by spoken words that were coherent (solid line) or anomalous in discourse (dotted line), for words in (*a*) sentence-medial and (*b*) sentence-final position, as well as for (*c*) short words (below 550 ms and a mean duration of 451 ms) and (*d*) long words (over 550 ms and a mean duration of 652 ms). The horizontal bars mark the range of acoustic word offsets for short and long CWs. (Reprinted with permission from Van Berkum *et al.* (2003). Copyright Elsevier B. V.)

owing to their frequent left anterior maximum over the scalp (Kluender & Kutas 1993). However, some particularly early left anterior negativities have also been claimed to directly reflect aspects of early syntactic processing (Friederici *et al.* 1996; see Friederici (1998) for a review). It is yet unclear as to what extent the early and later LAN effects reflect the same set of neuronal generators (see Friederici (1998), Hagoort *et al.* (1999), Brown & Hagoort (2000) and Brown *et al.* (2000) for discussions). The exact relationship between the LAN family of effects and the referentially induced sustained negative shift thus remains to be established.

In an additional study by Van Berkum *et al.* (2003), further evidence was obtained for the claim that the discourse context can affect the interpretation of an unfolding sentence extremely rapidly. This study compared discourse-dependent N400 effects with the standard N400 effect observed in response to a semantic anomaly. Subjects listened to short Dutch stories, in which the final sentence contained the CW. For each story, two CW alternatives were available: a discourse-coherent CW that was a good continuation of the earlier discourse and a discourse-anomalous CW that did not continue the discourse in a semantically plausible way (see the English translations of an example of the Dutch materials in (5); CWs are in bold). The difference in coherence between the two conditions hinged on considerable inferencing about the discourse topic and the situation it described. In the example, 'quick' and 'slow' are equally compatible with the sentence context, but only 'quick' is licensed by the discourse.

(5) As agreed upon, Jane was to wake her sister and her brother at five o'clock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally **quick/slow**.

When the final sentences were presented in isolation, no differences were obtained between the ERPs elicited by the CWs in the two conditions. However, in their discourse context, the discourse-anomalous CWs elicited a large and widely distributed negative deflection that emerged at approximately 150–200 ms after their acoustic onset, peaked at approximately 400 ms, lasted for approximately 800–1000 ms and reached its maximum over centroparietal electrode sites. In other words, a standard N400 effect was obtained that is indistinguishable from an N400 effect elicited by a semantic anomaly in a local sentence context (figure 7).

These findings show that the incoming words of an unfolding spoken sentence not only very rapidly make contact with 'global' discourse-level semantic information, but also do so in a way that is indistinguishable from how they make contact with 'local' sentence-level semantic information. The equivalent impact of sentence- and discourse-semantic contexts on the N400 elicited by a spoken word suggests that the process at work does not care where the semantic context originally came from, and evaluates the incoming words relative to the widest interpretive domain available. Moreover, these findings show that in natural spoken language comprehension, an unfolding word can be mapped onto discourse-level representations extremely rapidly, after only about three phonemes, and in many cases well before the end of the word.

5. CONCLUSIONS

In addition to discussing a number of relevant ERP/MEG studies on the nature of spoken language processing, some final remarks and conclusions will be made to highlight important aspects of electrical and magnetic measurements for this domain of research.

One particularly noteworthy aspect of a substantial number of ERP studies that were discussed above is that subjects were only engaged in the natural task, which in the case of speech is listening for understanding. An advantage of ERP/MEG recordings is that reliable effects can be obtained in the absence of potentially intrusive secondary tasks. Secondary tasks such as, for instance, lexical decision and word monitoring create a dual-task situation that might impact the primary process under investigation or the response characteristics (Norris *et al.* 2000). Getting reliable measurements in the absence of a secondary task is especially helpful in studies with language-impaired subjects, such as aphasic patients. For these patients, a dual-task situation might tax the processing resources beyond their limits. It is, nevertheless, reassuring that many of the ERP findings are consistent with the chronometric research on spoken word recognition and sentence processing.

Apart from the N200, all major effects discussed above are not only observed with spoken input, but also when subjects are reading. Since all these effects are triggered by processes beyond the single word level, such as integration, syntactic parsing or referential binding, this suggests that the functional architecture of language processing might be modality specific at the single word level, but beyond that spoken and written language comprehension seems to share processing principles and neural architecture. A recent series of ERP and fMRI studies on the integration of speech and co-speech gestures even suggests that similar principles apply beyond the domain of core language operations, with the left inferior frontal and the left superior temporal cortices as important nodes in the neural network for understanding (Hagoort 2005; Özyürek *et al.* 2007; Willems *et al.* 2007).

Even more remarkable is that, in general, the time course of the auditory N400 and P600/SPS (Hagoort & Brown 2000*a*,*b*), although often somewhat more extended, is similar compared with the N400 and P600/SPS elicited in reading. However, this similarity does not necessarily imply that the underlying processes have the same time course in both modalities. The crucial difference between reading and the processing of speech is the difference in the time at which word information is made available. In reading, words are essentially instantaneously available, whereas in speech the information accrues in a left-to-right order. Therefore, relative to the availability of information in reading, the N400 and P600/SPS effects as well as the referential negativity in speech are actually remarkably early. The time course of the ERP effects is compatible with the view that the different information types (lexical, syntactic, phonological, pragmatic) are processed in parallel and influence the interpretation process incrementally, i.e. as soon as the relevant pieces of information are available (Marslen-Wilson 1989; Zwitserlood 1989; Jackendoff 2002). I refer to this as the immediacy principle.

Furthermore, the ERP results clearly show that there does not seem to be a separate stage during which word meaning is exclusively integrated at the sentence level. Incremental interpretation is, for the most part, done by an immediate mapping onto a discourse model. Moreover, listeners relate the incoming acoustic signal to the semantic context not just before they have heard the complete word but before they can actually know exactly what the unfolding word itself is going to be. The presented ERP evidence is consistent with chronometric data, in showing that discourse- and sentence-semantic information differentially affects the processing of word candidates before the acoustic input itself uniquely specifies a word.

Finally, a cognitive neuroscience account of human language processing should not only specify the crucial network of brain areas supporting listening, reading and speaking. In addition, the temporal dynamics of the neurophysiological processes that instantiate these language functions has to be specified as well. EEG and MEG recordings are superior over any other non-invasive method in providing a window onto exactly these dynamics. As I have shown here, the limited number of studies on speech that are currently available already have provided interesting insights into the organization of spoken language understanding.

REFERENCES

Ainsworth-Darnell, K., Shulman, H. & Boland, J. 1998 Dissociating brain responses to syntactic and semantic anomalies: evidence from event-related potentials. *J. Mem. Lang.* **38**, 112–130. (doi:10.1006/jmla.1997.2537)

Altmann, G. T. M. 1988 Ambiguity, parsing strategies, and computational models. *Lang. Cogn. Process.* **3**, 73–97.

Boland, J. E. & Cutler, A. 1996 Interaction with autonomy: multiple output models and the inadequacy of the great divide. *Cognition* **58**, 309–320. (doi:10.1016/0010-0277(95)00684-2)

Brown, C. M. & Hagoort, P. 1993 The processing nature of the N400: evidence from masked priming. *J. Cogn. Neurosci.* **5**, 34–44.

Brown, C. M. & Hagoort, P. 2000 On the electrophysiology of language comprehension: implications for the human language system. In *Architectures and mechanisms for language processing* (eds M. W. Crocker, M. Pickering & C. Clifton), pp. 213–237. Cambridge, UK: Cambridge University Press.

Brown, C. M., Hagoort, P. & Chwilla, D. J. 2000 An event-related brain potential analysis of visual word priming effects. *Brain Lang.* **72**, 158–190. (doi:10.1006/brln.1999.2284)

Catani, M., Jones, D. K. & Ffytche, D. H. 2004 Perisylvian language networks of the human brain. *Ann. Neurol.* **57**, 8–16. (doi:10.1002/ana.20319)

Connolly, J. F. & Phillips, N. A. 1994 Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. *J. Cogn. Neurosci.* **6**, 256–266.

Connolly, J. F., Stewart, S. H. & Phillips, N. A. 1990 The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain Lang.* **39**, 302–318. (doi:10.1016/0093-934X(90)90016-A)

Connolly, J. F., Phillips, N. A., Stewart, S. H. & Brake, W. G. 1992 Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain Lang.* **43**, 1–18. (doi:10.1016/0093-934X(92)90018-A)

Connolly, J. F., Phillips, N. A. & Forbes, K. A. 1995 The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalogr. Clin. Neuropsychol.* **94**, 276–287. (doi:10.1016/0013-4694(95)98479-R)

Coulson, S., King, J. W. & Kutas, M. 1998 Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* **13**, 21–58. (doi:10.1080/016909698386582)

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.

DeLong, K. A., Urbach, T. P. & Kutas, M. 2005 Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* **8**, 1117–1121. (doi:10.1038/nn1504)

Donaldson, D. I. & Rugg, M. D. 1999 Event-related potential studies of associative recognition and recall: electrophysiological evidence for context dependent retrieval processes. *Brain Res. Cogn. Brain Res.* **8**, 1–16. (doi:10.1016/S0926-6410(98)00051-2)

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M. & Marantz, A. 2001 A magnetoencephalographic component whose latency reflects lexical frequency. *Brain Res. Cogn. Brain Res.* **10**, 345–348. (doi:10.1016/S0926-6410(00)00053-7)

Ferreira, F. & Clifton, C. J. 1986 The independence of syntactic processing. *J. Mem. Lang.* **25**, 348–368. (doi:10.1016/0749-596X(86)90006-9)

Fiebach, C. J., Schlesewsky, M. & Friederici, A. D. 2001 Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. *J. Psycholinguist. Res.* **30**, 321–338. (doi:10.1023/A:1010447102554)

Fodor, J. D. 1983 *The modularity of mind*. Cambridge, MA: MIT Press.

Frauenfelder, U. H. & Tyler, L. K. 1987 The process of spoken word recognition: an introduction. *Cognition* **25**, 1–20. (doi:10.1016/0010-0277(87)90002-3)

Frazier, L. 1987 Sentence processing: a tutorial review. In *Attention and performance XII* (ed. M. Coltheart), pp. 559–585. Hillsdale, NJ: Lawrence Erlbaum.

Friederici, A. D. 1995 The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain Lang.* **50**, 259–281. (doi:10.1006/brln.1995.1048)

Friederici, A. D. 1998 The neurobiology of language comprehension. In *Language comprehension: a biological perspective* (ed. A. D. Friederici), pp. 263–301. Berlin, Germany: Springer.

Friederici, A. D. 2002 Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* **6**, 78–84. (doi:10.1016/S1364-6613(00)01839-8)

Friederici, A. D., Pfeifer, E. & Hahne, A. 1993 Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* **1**, 183–192. (doi:10.1016/0926-6410(93)90026-2)

Friederici, A. D., Hahne, A. & Mecklinger, A. 1996 Temporal structure of syntactic parsing: early and late event-related brain potential effects. *J. Exp. Psychol. Learn.* **22**, 1219–1248. (doi:10.1037/0278-7393.22.5.1219)

Gibson, E. 1998 Linguistic comlexity: locality of syntactic dependencies. *Cognition* **68**, 1–76. (doi:10.1016/S0010-0277(98)00034-1)

Grosjean, F. 1980 Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* **28**, 267–283.

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: Erlbaum.

Hagoort, P. & Brown, C. M. 2000*a* ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia* **38**, 1518–1530. (doi:10.1016/S0028-3932(00)00052-X)

Hagoort, P. & Brown, C. M. 2000*b* ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* **38**, 1531–1549. (doi:10.1016/S0028-3932(00)00053-1)

Hagoort, P., Brown, C. M. & Groothusen, J. 1993 The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* **8**, 439–483.

Hagoort, P., Brown, C. M. & Osterhout, L. 1999 The neurocognition of syntactic processing. In *Neurocognition of language* (eds C. M. Brown & P. Hagoort), pp. 273–317. Oxford, UK: Oxford University Press.

Hagoort, P., Wassenaar, M. & Brown, C. M. 2003 Syntax-related ERP-effects in Dutch. *Cogn. Brain Res.* **16**, 38–50. (doi:10.1016/S0926-6410(02)00208-2)

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)

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D. & Dale, A. M. 2002 N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage* **17**, 1101–1116. (doi:10.1006/nimg.2002.1268)

Helenius, P., Salmelin, R., Service, E. & Connolly, J. F. 1998 Distinct time courses of word and context comprehension in the left temporal cortex. *Brain* **121**, 1133–1142. (doi:10.1093/brain/121.6.1133)

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S. & Lyytinen, H. 2002 Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J. Neurosci.* **22**, 2936–2944.

Holcomb, P. J. 1993 Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* **30**, 47–61.

Holcomb, P. J. & Neville, H. J. 1990 Semantic priming in visual and auditory lexical decision: a between modality comparison. *Lang. Cogn. Process.* **5**, 281–312.

Holcomb, P. J. & Neville, H. J. 1991 Natural speech processing: an analysis using event-related brain potentials. *Psychobiology* **19**, 286–300.

Jackendoff, R. 1999 The representational structures of the language faculty and their interactions. In *The neurocognition of language* (eds C. M. Brown & P. Hagoort), pp. 37–79. Oxford, UK: Oxford University Press.

Jackendoff, R. 2002 *Foundations of language: brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jensen, O., Weder, N., Bastiaansen, M., van den Brink, D., Dijkstra, T. & Hagoort, P. In preparation. Modulation of the beta rhythm in a language comprehension task.

Kaan, E. & Swaab, T. Y. 2003 Repair, revision and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* **15**, 98–110. (doi:10.1162/089892903321107855)

Kaan, E., Harris, A., Gibson, E. & Holcomb, P. J. 2000 The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* **15**, 159–201. (doi:10.1080/016909600386084)

Kellenbach, M. I. & Michie, P. T. 1996 Modulation of event-related potentials by semantic priming: effects of color-dued selective attention. *J. Cogn. Neurosci.* **8**, 155–173.

Kim, A. & Osterhout, L. 2005 The independence of combinatory semantic processing: evidence from event-related potentials. *J. Mem. Lang.* **52**, 205–225. (doi:10.1016/j.jml.2004.10.002)

King, J. W. & Kutas, M. 1995 Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* **7**, 376–395.

Kintsch, W. 1998 *Comprehension: a paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Kluender, R. & Kutas, M. 1993 Subjacency as a processing phenomenon. *Lang. Cogn. Process.* **8**, 573–633.
Kuperberg, G. R., Sitnikova, T., Caplan, D. & Holcomb, P. J. 2003 Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cogn. Brain Res.* **17**, 117–129. (doi:10.1016/S0926-6410(03)00086-7)
Kutas, M. 1997 Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology* **34**, 383–398. (doi:10.1111/j.1469-8986.1997.tb02382.x)
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. New York, NY: Academic Press.
Kutas, M., Federmeier, K. D., Coulson, S., King, J. W. & Münte, T. F. 2000 Language. In *Handbook of psychophysiology* (eds J. T. Cacioppo, L. G. Tassinary & G. G. Berntson), pp. 576–601. New York, NY: Cambridge University Press.
Luck, S. J. 2005 *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.
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)
Marslen-Wilson, W. D. 1973 Linguistic structure and speech shadowing at very short latencies. *Nature* **244**, 522–523. (doi:10.1038/244522a0)
Marslen-Wilson, W. D. 1984 Function and process in spoken word-recognition. In *Attention and performance X* (eds H. Bouma & D. Bouwhuis), pp. 125–150. Hillsdale, NJ: Erlbaum.
Marslen-Wilson, W. D. 1987 Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102. (doi:10.1016/0010-0277(87)90005-9)
Marslen-Wilson, W. D. 1989 Access and integration: projecting sound onto meaning. In *Lexical representation and process* (ed. W. D. Marslen-Wilson), pp. 3–24. Cambridge, MA: MIT Press.
Marslen-Wilson, W. D. 1993 Issues of process and representation in lexical access. In *Cognitive models of speech processing: the second sperlonga meeting* (eds G. T. M. Altmann & R. Shillock), pp. 187–210. Hillsdale, NJ: Erlbaum.
Marslen-Wilson, W. D. & Tyler, L. K. 1980 The temporal structure of spoken language understanding. *Cognition* **8**, 1–71. (doi:10.1016/0010-0277(80)90015-3)
Marslen-Wilson, W. D. & Welsh, A. 1978 Processing interactions and lexical access during word-recognition in continuous speech. *Cogn. Psychol* **10**, 29–63. (doi:10.1016/0010-0285(78)90018-X)
McCallum, W. C., Farmer, S. F. & Pocock, P. V. 1984 The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalogr. Clin. Neuropsychol.* **59**, 477–488. (doi:10.1016/0168-5597(84)90006-6)
McClelland, J. L. & Elman, J. L. 1986 The TRACE model of speech perception. *Cogn. Psychol.* **18**, 1–86. (doi:10.1016/0010-0285(86)90015-0)
McKinnon, R. & Osterhout, L. 1996 Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. *Lang. Cogn. Process.* **11**, 495–523. (doi:10.1080/016909696387132)
Mitchell, D. C., Cuetos, F., Corley, M. M. B. & Brysbaert, M. 1995 Exposure-based models of human parsing: evidence for the use of coarse-grained (nonlexical) statistical records. *J. Psycholinguist. Res.* **24**, 469–488. (doi:10.1007/BF02143162)
Morton, J. 1969 Interaction of information in word recognition. *Psychol. Rev.* **76**, 165–178. (doi:10.1037/h0027366)
Münte, T. F. & Heinze, H. J. 1994 ERP negativities during syntactic processing of written words. In *Cognitive electrophysiology* (eds H. J. Heinze, T. F. Münte & H. R. Mangun), pp. 211–238. Boston, MA: Birkhauser.
Münte, T. F., Heinze, H. J. & Mangun, G. R. 1993 Dissociation of brain activity related to syntactic and semantic aspects of language. *J. Cogn. Neurosci.* **5**, 335–344.
Münte, T. F., Matzke, M. & Johannes, S. 1997 Brain activity associated with syntactic incongruities in words and pseudo-words. *J. Cogn. Neurosci.* **9**, 300–311.
Münte, T. F., Schiltz, K. & Kutas, M. 1998 When temporal terms belie conceptual order. *Nature* **395**, 71–73. (doi:10.1038/25731)
Myers, J. L. & O'Brien, E. J. 1998 Accessing the discourse representation during reading. *Discourse Process.* **26**, 131–157.
Näätänen, R. 2001 The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* **38**, 1–21.
Näätänen, R. *et al.* 1997 Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432–434. (doi:10.1038/385432a0)
Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I. & Garrett, M. F. 1991 Syntactically based sentence processing classes: evidence from event-related brain potentials. *J. Cogn. Neurosci.* **3**, 151–165.
Norris, D. 1994 Shortlist: a connectionist model of continuous speech recognition. *Cognition* **52**, 189–234. (doi:10.1016/0010-0277(94)90043-4)
Norris, D., McQueen, J. M. & Cutler, A. 2000 Merging information in speech recognition: feedback is never necessary. *Behav. Brain Sci.* **23**, 299–370. (doi:10.1017/S0140525X00003241)
O'Rourke, T. B. & Holcomb, P. J. 2002 Electrophysiological evidence for the efficiency of spoken word processing. *Biol. Psychol.* **60**, 121–150. (doi:10.1016/S0301-0511(02)00045-5)
Osterhout, L. 1997 On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang.* **59**, 494–522. (doi:10.1006/brln.1997.1793)
Osterhout, L. & Hagoort, P. 1999 A superficial resemblance doesn't necessarily mean you're part of the family: counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS debate. *Lang. Cogn. Process.* **14**, 1–14. (doi:10.1080/016909699386356)
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. 1993 Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Lang. Cogn. Process.* **8**, 413–438.
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. & Mobley, L. A. 1995 Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* **34**, 739–773. (doi:10.1006/jmla.1995.1033)
Osterhout, L., Holcomb, P. J. & Swinney, D. A. 1994 Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exp. Psychol. Learn.* **20**, 786–803. (doi:10.1037/0278-7393.20.4.786)

Osterhout, L., Bersick, M. & McLaughlin, J. 1997 Brain potentials reflect violations of gender stereotypes. *Mem. Cognition* **25**, 273–285.

Özyürek, A., Willems, R. M., Kita, S. & Hagoort, P. 2007 On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* **4**, 605–616. (doi:10.1162/089892903322370807)

Pulvermüller, F., Shtyrov, Y., Kujala, T. & Näätänen, R. 2004 Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology* **41**, 106–112. (doi:10.1111/j.1469-8986.2003.00135.x)

Rösler, F., Friederici, A. D., Pütz, P. & Hahne, A. 1993 Event-related brain potentials while encountering semantic and syntactic constraint violations. *J. Cogn. Neurosci.* **5**, 345–362.

Rugg, M. D. & Allan, K. 2000 Event-related potential studies of long-term memory. In *The Oxford handbook of memory* (eds E. Tulving & F. I. M. Craik), pp. 521–537. Oxford, UK: Oxford University Press.

Schmitt, B. M., Lamers, M. J. A. & Münte, T. F. 2002 Electrophysiological estimates of biological and syntactic gender violation during pronoun. *Cogn. Brain Res.* **14**, 333–346. (doi:10.1016/S0926-6410(02)00136-2)

Streb, J., Rösler, F. & Hennighausen, E. 1999 Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain Lang.* **70**, 273–286. (doi:10.1006/brln.1999.2177)

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.

Tyler, L. K. & Marslen-Wilson, W. D. 1977 The on-line effects of semantic context on syntactic processing. *J. Verbal Learn. Verbal Behav.* **16**, 683–692. (doi:10.1016/S0022-5371(77)80027-3)

Tyler, L. K. & Wessels, J. 1983 Quantifying contextual contributions to word-recognition processes. *Percept. Psychophys.* **34**, 409–420.

Van Berkum, J. J. A., Brown, C. M. & Hagoort, P. 1999*a* Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* **41**, 147–182. (doi:10.1006/jmla.1999.2641)

Van Berkum, J. J. A., Brown, C. M. & Hagoort, P. 1999*b* When does gender constrain parsing? Evidence from ERPs. *J. Psycholinguist. Res.* **28**, 555–566. (doi:10.1023/A:1023224628266)

Van Berkum, J. J. A., Brown, C. M., Hagoort, P. & Zwitserlood, P. 2003 Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology* **40**, 235–248. (doi:10.1111/1469-8986.00025)

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 den Brink, D., Brown, C. M. & Hagoort, P. 2001 Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn. Neurosci.* **13**, 967–985. (doi:10.1162/089892901753165872)

Van den Brink, D., Brown, C. M. & Hagoort, P. 2006 The cascaded nature of lexical selection and integration in auditory sentence processing. *J. Exp. Psychol. Learn.* **32**, 364–372. (doi:10.1037/0278-7393.32.3.364)

Van Herten, M., Kolk, H. H. J. & Chwilla, D. J. 2004 An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* **22**, 241–255. (doi:10.1016/j.cogbrainres.2004.09.002)

Van Petten, C., Coulson, S., Rubin, S., Plante, E. & Parks, M. 1999 Time course of word identification and semantic integration in spoken language. *J. Exp. Psychol. Learn.* **25**, 394–417. (doi:10.1037/0278-7393.25.2.394)

Vos, S. H., Gunter, T. C., Kolk, H. H. & Mulder, G. 2001 Working memory constraints on syntactic processing: an electrophysiological investigation. *Psychophysiology* **38**, 41–63. (doi:10.1111/1469-8986.3810041)

Wassenaar, M. & Hagoort, P. 2007 Thematic role assignment in patients with Broca's aphasia: sentence–picture matching electrified. *Neuropsychologia.* **45**, 716–740. (doi:10.1016/j.neuropsychologia.2006.08.016)

Wicha, N. Y. Y., Moreno, E. M. & Kutas, M. 2004 Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* **16**, 1272–1288. (doi:10.1162/0898929041920487)

Willems, R. M., Özyürek, A. & Hagoort, P. 2007 When language meets action: the neural integration of gesture and speech. *Cerebral Cortex* **17**, 2322–2333. (doi:10.1093/cercor/bhL141)

Zwitserlood, P. 1989 The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* **32**, 25–64. (doi:10.1016/0010-0277(89)90013-9)