



ERP effects of listening to speech: semantic ERP effects

Peter Hagoort*, Colin M. Brown

Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

Received 29 January 1997; received in revised form 10 January 2000; accepted 11 February 2000

Abstract

In this study, event-related brain potential effects of speech processing are obtained and compared to similar effects in sentence reading. In two experiments spoken sentences were presented with semantic violations in sentence-final or mid-sentence positions. For these violations N400 effects were obtained that were very similar to N400 effects obtained in reading. However, the N400 effects in speech were preceded by an earlier negativity (N250). This negativity is not commonly observed with written input. The early effect is explained as a manifestation of a mismatch between the word forms expected on the basis of the context, and the actual cohort of activated word candidates that is generated on the basis of the speech signal. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: N400; N250; Natural speech; Electrophysiology; Lexical selection; Lexical integration

1. Introduction

The evolutionary hardwired system for language mainly operates through speaking and listening. Reading and writing are derived from our competence in understanding and producing speech, by mapping speech sounds onto a set of conventional symbols agreed upon by the language community that uses a particular reading and writing system. Notwithstanding the primacy of speech as the only common mode of language processing in the human species, it is reading that has been investigated most extensively in psycholinguistics. Recent brain imaging studies of language functions (including techniques such as ERP, MEG, PET, fMRI) have also more often used written than spoken input. However, a full understanding of the human language faculty requires that, in addition, we study the processing of speech directly. In this study, we primarily investigate electrophysiological correlates of processes involved in spoken sentence

processing. In addition, we will compare effects of listening to language with those of reading.

It is generally assumed that in reading and listening, apart from the modality dependent form-level representations, the same knowledge sources are exploited in mapping the input onto the meaning of the whole utterance. That is, under both input conditions syntactic and semantic information associated with lexical entries, and the knowledge of the grammar required for assigning structure to the incoming strings of words, is assumed to be the same. However, the nature of the input might impose substantially different constraints on the time course and the computational characteristics of accessing and exploiting these knowledge sources in real time.

In this study, we report Event Related Potential (ERP) effects to spoken sentences. ERPs reflect the sum of simultaneous post-synaptic activity of many neurons, recorded at the scalp as small voltage fluctuations in the EEG time locked to sensory, motor, or cognitive processes. Our aim here is to determine in more detail the characteristics of one particular ERP effect in relation to spoken input. This is the so-called N400 effect.

Broadly speaking, the N400 is related to the mean-

* Corresponding author. Fax: +31-24-352-1213.
E-mail address: peha@mpi.nl (P. Hagoort).

ing of the sentence [22]. One way in which N400 effects can be elicited is by a semantic mismatch between the meaning of a word and the semantic specifications of its sentence context (e.g., “He spread the warm bread with socks.” vs “He spread the warm bread with butter.”). Averaged over a number of trials, a substantial negative deflection is visible in the ERP to semantically incongruous words (i.e. *socks*) relative to the wave-shape elicited by congruous words such as *butter*. The difference in the amplitude of this negative deflection as a result of the difference in semantic fit between a word and its context is known as the N400 effect.

To date, the majority of N400 studies on sentence processing focused on N400 effects to written input. The general finding of these studies is that ERPs to words that were either semantically anomalous or relatively unexpected in their sentence context, showed an increased negative deflection. This increase was seen relative to ERPs elicited by words that fit their context well ([22]; see for an overview, [24, 35]). In general, meaningful words elicit the N400 component. In lexical and sentential contexts the amplitude of this component is modulated by the preceding semantic information. Since its first report [22], the differential monophasic negative shift is known as the N400 effect, and it has been shown to be a sensitive measure of on-line semantic analysis. The by now classical N400 effect has a posterior distribution, with an onset around 200–250 ms after word presentation, and with a maximal amplitude between 380 and 440 ms. In response to written words, the N400 effect is usually (but not always) slightly larger over the right than the left hemisphere, and has a duration of between 300 and 400 ms in the averaged waveform [24].

Relatively few studies have focused on N400 effects in spoken sentences [1, 5, 6, 8, 9, 11, 14, 17, 18, 30, 32, 36, 37, 39]. Some of these studies do not provide a fully accurate view on ERPs to spoken sentence processing because the speech signal clearly deviated from natural speech. These deviations resulted from using synthetic speech and adding artificial pauses between words [1, 14], from speaking sentences at a very slow speaking rate [18], by splicing the critical sentence-final words from other sentences into the experimental sentences [30], or by separating the sentence context and the sentence-final target word by a pause [9, 37, 39].

In all the studies that used natural connected speech, N400 effects were only investigated for spoken words in sentence-final position. Overall, for N400 effects in sentence-final position differences between visual and auditory N400 effects are reported with respect to onset latency, duration, and scalp distribution. Holcomb and Neville [16] reported a very early effect to sentence-final 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. In contrast, the onset of the auditory N400 effect in the Friederici et al. study [11] was in the same range as the onset of the visual N400 effect. In addition, the auditory N400 effects in both the McCallum et al. and in the Holcomb and Neville studies [30, 16] were more sustained over frontal than over posterior sites. On the whole, auditory N400 effects are of longer duration than their visual counterparts.

Especially the earlier onset of N400 effects in connected speech is surprising. In contrast to written words, spoken words are encoded in a signal that is extended over time. Mapping the spoken signal onto information in the mental lexicon is therefore presumably in many ways quite 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 [29]. 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 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. For instance, when presented in isolation, the recognition point 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 recognition point is located well before the end of the word [25, 26]. It has been claimed that context information can be used to shift the recognition point to an earlier part of the speech signal [40]. 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 [12, 26].

Selection of the appropriate word form is one step in the process of understanding the content of an utterance. The mental lexicon is the crucial interface between language form and language content, two fundamentally distinct knowledge domains [27]. Lexical access refers to the mapping of the input signal onto word form representations in the mental lexicon. This mapping results in the retrieval of information associated with the word's form, including its syntactic properties (e.g., gender, word class) and meaning. However, since we normally perceive words not in isolation, but in the context of other words, the retrieved syntactic and semantic information has to be integrated with the higher level context representation of the preceding utterance. We will refer to this process as lexical integration.

N400 effects are claimed to be especially sensitive to lexical integration; that is, to matching the content spe-

cifications of a lexical candidate against the content specifications of the word or sentence context [4, 15, 21]. Although the precise time course of lexical access and lexical integration is still a matter of dispute [10, 27], an N400 integration effect starting at 50–100 ms after the onset of spoken words that had an average duration of 516 ms in the Holcomb and Neville [17] study would be remarkably early on anybody's account of spoken word processing, even when coarticulation effects are taken into consideration.

To claim evidence for an N400 effect with a substantially earlier onset in speech than in reading presupposes that for both language modalities there is a clearly similar monophasic negative shift in the ERP waveforms. However, there is some doubt as to whether the auditory N400 effect is indeed a similar monophasic negative shift as the visual N400 effect. 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. McCallum et al. [30] 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 and colleagues [5, 6, 8]. These authors compared ERPs to sentence-final words in highly constraining sentence contexts (e.g., "The king wore a golden crown.") with sentence-final ERPs in sentence contexts with low constraints (e.g., "The woman talked about the frogs.") [6, 8]. A negative shift was observed to final words of low constraining sentences relative to final words in highly constraining sentences. The authors report [6] that individual difference waveforms (but not the grand averages), obtained by subtracting for each subject from each other the waveforms to the sentence-final words 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). Moreover, adding an auditory mask to the sentence-final word resulted in a latency shift of the N400 peak, but did not affect N200 latencies [8]. Again, this could not be observed in the average waveforms, but only by determining the peak latencies for individual subjects across sites within latency ranges of 150–350 ms and 350–600 ms respectively. The authors suggest that the N200 effects reflect acoustic/phonological processing of the sentence-final word. That is, if the initial phoneme of the sentence-final word mismatches with the onset of the expected word, an N200 effect emerges. The N400 amplitude is claimed to be modulated by semantic expectancy. However, since both effects depend on a vio-

lation of semantic expectancy, it is not immediately clear how functionally different their accounts for these two ERP effects really are (for a more explicit account of the processing difference underlying the early negativity versus the N400 effect, see below). Moreover, the absence of a clearly visible N200 effect in the averaged waveform is of some concern, since determining peaks that can be unambiguously assigned to a particular component in individual subject waveforms is not always easy for componentry in the specified latency ranges, as becomes clear on inspection of the individual waveforms that are presented in these studies.

Recently, Connolly and Phillips [5] 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 sentence-final 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 [28] or a shortlist [31] 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 sentence-final 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.

It is interesting to note that Connolly, Phillips and Forbes [7] repeated their 1994 study with written presentation. The same sentences presented in the visual modality resulted in N400 effects, but a similar early effect as in the auditory modality did not appear in the visual modality. But, again, detecting two clearly distinct peaks in the averaged waveforms of their 1994 study is not all that easy.

In short, the earliness of N400 effects in connected speech might in fact be due to another effect around 200 ms post-onset. Inspection of the waveforms in some of the N400 studies with spoken sentences indeed suggests the presence of an early negative shift that is separate from the following N400 effect [17, 18]. The same suggestion of a biphasic effect is clearly present in a study that presented spoken words in the context of one other spoken (prime) word [17]. The most explicit account of this early negative effect has been provided by Connolly and colleagues. However, in their studies this early negative effect is not always clearly visible in the averaged waveforms.

A recent study by Van Petten and colleagues [37] did not substantiate the separation of the overall negative shift into an early negativity followed by a later N400 effect. 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*), (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 to 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.

Given some of the inconsistencies in the results reported in the literature, the ERP effects of sentence processing in spoken language need replication and further investigation. In this study, we decided to further investigate the nature of semantic anomaly effects in ERPs to spoken words in naturally produced connected speech. In order to investigate the generality of the effects for the on-line processing of speech, we tested semantic anomaly effects not only in sentence-final, but also in sentence-medial positions. To date, N400 effects in connected speech have only been tested and observed for words in sentence-final position. The clear prosodic marking of the end of the sentence potentially leads to sentence wrap-up effects that might strengthen, reduce, or otherwise modify the effects seen in other positions of the sentence. It is therefore im-

portant to obtain auditory N400 effects that are not contaminated by sentence closure effects. A first aim of this study, then, was to see whether similar N400 effects are observed in different positions in the sentence, with concomitant differences in prosodic marking. In addition, we were especially interested to see whether in both sentence-medial and sentence-final positions the semantic anomalies led to monophasic N400 effects known from the visual modality, or to a biphasic negativity composed of an earlier part and an N400 effect.

A particularly noteworthy aspect of our study is that subjects were only engaged in the natural task, which in the case of speech is listening for understanding. An advantage of ERP recordings is that reliable effects can be obtained in the absence of potentially intrusive secondary tasks. Secondary tasks such as, for instance, lexical decision or word monitoring create a dual-task situation which might impact the primary process under investigation. Getting reliable measurements in the absence of a secondary task is especially helpful in studies with language-impaired subjects.

In summary, this study is different from all previous studies on N400 effects in spoken sentences in combining the presentation of sentences spoken at a natural rate without artificial pauses with semantic violations in sentence-final as well as mid-sentence positions under the natural task condition of listening to speech for understanding.

2. Experiment 1: sentence-final semantic anomalies

2.1. Method

2.1.1. Materials

The experimental items consisted of a set of 120 Dutch sentences. Half of these sentences ended with a word that matched the sentential-semantic constraints, e.g., “The children like to play in the *garden*.” The other half of the sentences ended in an anomalous way with a word that violated the sentential-semantic constraints, e.g., “The girl dropped the candy on the *sky*.” (The full set of materials is available upon request.)

The 120 critical sentences were derived from a larger set of sentence materials that had been submitted to a cloze test. In each test, 24 subjects who did not participate in the subsequent ERP studies were given a list of written sentences with the final word omitted. Subjects were asked to read the sentences and to fill in the first word that came to mind at the position of the omitted sentence-final word. On the basis of the cloze test results, a set of 120 sentences was selected with a cloze probability higher than 29%.

Sixty of the 120 selected sentences were assigned to the congruent condition. These sentences had a mean

cloze probability of 60.3% (range=29.9–100). The remaining set of 60 sentences was used to construct the anomalous items. These sentences had a mean cloze probability of 54.6% (range=29.2–100). In this way the sentence contexts in both conditions were approximately equally constraining for the sentence-final word. The syntactic structure of the sentences was matched between the two conditions. The mean sentence length was 7.8 words in both conditions (range=5–10 words).

The items for the anomalous condition were constructed by replacing the sentence-final words in the 60 original sentences by words that did not fit the semantic constraints of the preceding context. With the exception of four items, the anomalous words did not share any word-initial sounds with the original congruent or best completions.

All sentence-final words were nouns with a mean duration of 489 ms (SD=81 ms) in the congruent condition and 544 ms (SD=95 ms) in the anomalous condition. Each sentence-final word began with a plosive. In this way the last word in the continuous speech signal had a clear physical marker that facilitated the alignment of the ERP waveform to the onset of the sentence-final word. The words preceding the sentence-final words were similar in length in the congruent (mean length=3.5 phonemes; SD=1.8) and anomalous condition (mean length=2.8 phonemes; SD=1.7). In both conditions, the majority of words directly preceding the critical word in sentence-final position were articles (in Dutch: *de, het, een*) or possessive pronouns (in Dutch: *mijn, zijn, haar*). The lexical characteristics of the sentence pre-final words were therefore very similar in both conditions.

In addition to the set of 120 experimental items, a set of 90 filler items was constructed. Half of these filler sentences contained semantic anomalies in different non-final positions. These materials were included to prevent subjects from developing a strategy of predicting the position of the anomalous word. Finally, the experimental list was preceded by a practice list of 20 items, ten of which contained a semantic anomaly in a non-final position.

The sentences were all spoken by an experienced female speaker, with normal intonation at a normal speaking rate, and read into the computer with a sampling rate of 20 kHz. A speech waveform editing system was used to mark the onset of the sentence-final words. The interval of silence between sentences ranged between 4 and 5.5 s. After randomizing the sentences, a digital audio tape was recorded containing the experimental list preceded by the practice items.

2.1.2. Subjects

None of the subjects in this and the following experiment had any neurological impairment or had experi-

enced any neurological trauma according to their responses on a questionnaire. The handedness of the subjects was established via an adapted Dutch version of the Oldfield Handedness Inventory [33]. The subjects were paid for their participation. Twelve subjects (10 female; eight right-handed) participated in Experiment 1.

2.1.3. Procedure

The subjects were tested individually in a dimly illuminated sound-attenuating booth. Subjects were seated in a comfortable reclining chair, instructed to move as little as possible, and to keep their eyes fixated on a point at eye-level 1.5 m in front of them. They were told that they would hear sentences that they should listen to for understanding and with full attention. In addition, they were told that some sentences had an odd word in them. No additional task demands were imposed. All the stimuli were presented via a DAT-recorder (SONY 300ES). The subjects listened to the stimuli via a closed-ear Sennheiser HMD-224 headphone.

2.1.4. EEG-recording

The EEG was recorded from seven tin electrodes in an electrode cap. Each electrode was referred to the left mastoid. Three electrodes were placed according to the international 10–20 system [20] over midline sites at Fz, Cz, and Pz locations. Four electrodes were placed laterally over symmetrical anterior and posterior positions. The anterior temporal electrodes (ATL, ATR) were located one half of the distance between F7 and T3, and F8 and T4 respectively. Symmetrical posterior temporal electrodes (PTL, PTR) were placed lateral, by 30% of the interaural distance, and 13% posterior to the vertex. Vertical eye movements and blinks were monitored via a supra- to sub-orbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye-movements. Activity over the right mastoid bone was recorded on an additional channel to determine if there were differential contributions of the experimental variables to the two presumably neutral mastoid sites. No such differential effects were observed.

The EEG and EOG recordings were amplified with Nihon Kohden AB-601G bio-electric amplifiers, using a Hi-Cut of 30 Hz and a time constant of 8 s. Impedances were kept below 5 k Ω . The EEG and EOG were digitized on-line with a sampling frequency of 200 Hz. Sampling started 150 ms before the presentation of the sentences and continued for a period of 5.5 s. The experimental trials were stored along with condition codes for off-line averaging and data analysis.

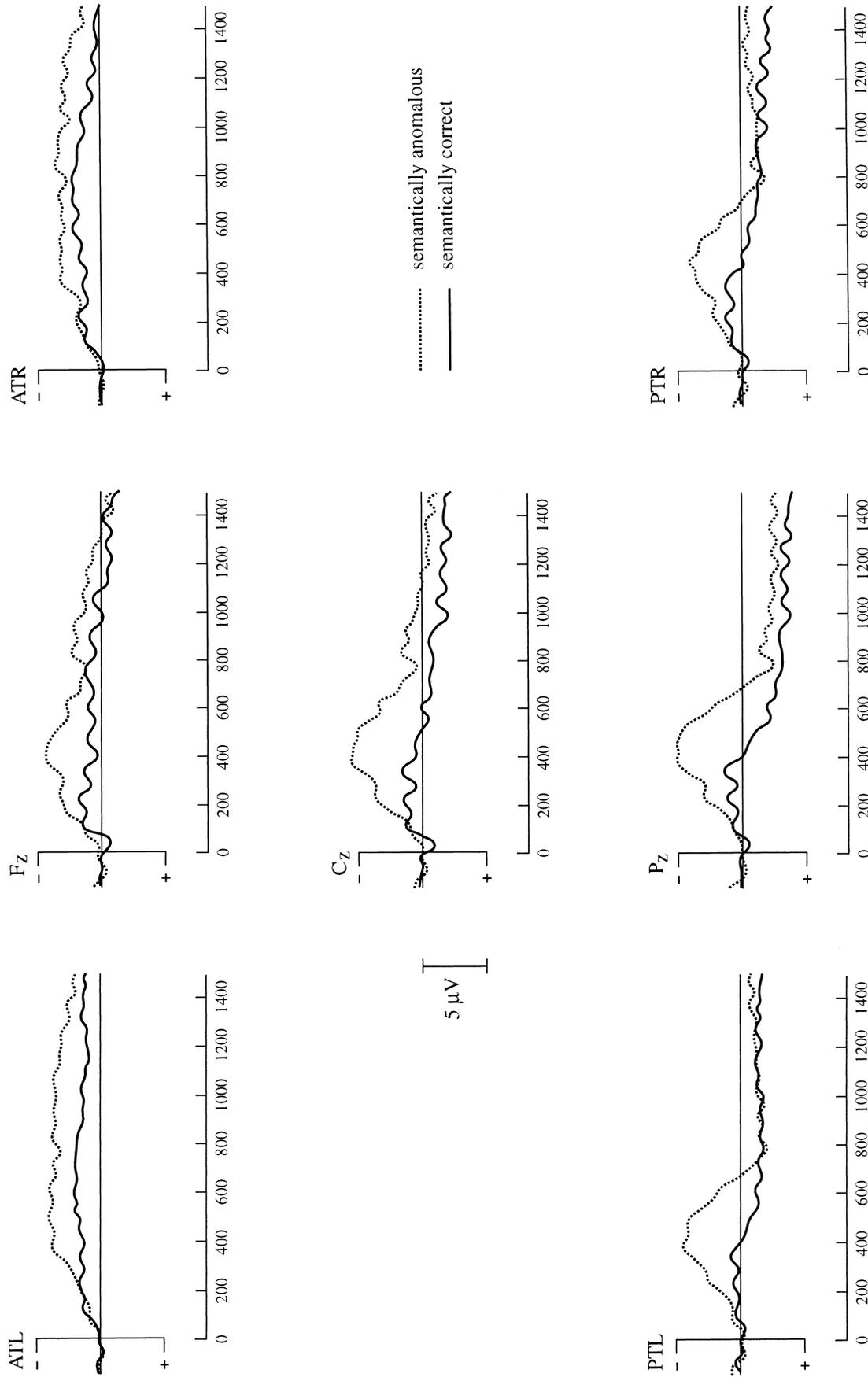


Fig. 1. Connected speech. Grand average ERPs (across all subjects and items), to the sentence-final semantic anomalies, and to the correct sentence endings. The onset of the sentence-final words is at zero. The time axis is in milliseconds. Each plot represents approximately 628 trials. Negative polarity is plotted upwards in this and subsequent figures.

2.2. Results

Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking and EMG artifacts. Trials containing such artifacts were rejected. For the critical latency window around the sentence-final word, the overall rejection rate was 12.7%.

Fig. 1 shows the grand average waveforms by electrode site for the sentence-final congruent and the sentence-final anomalous words. The first thing to note is that compared to the reading of sentences [23], the early N1 and P2 components are very much reduced. The reduction or absence in the waveforms of these components is most likely due to the continuous nature of the speech signal, in which clear physical boundaries between the words of the sentences are lacking. Although we perceive words as separate entities in spoken sentences, the speech waveform of these sentences often shows no clearly identifiable word boundaries. With respect to the ERP, either the continuous speech signal does not seem to have clear cues that elicit the N1 and P2 at word boundary positions, or the silence preceding the closure release of the word-initial stop consonants is too short to span the refractory period of these components.

The most salient feature of the waveforms in Fig. 1 is the large negativity to the anomalous sentence endings. This negativity is composed of an early part, and a later part. The early part shows a separation of the waveforms for the anomalous and congruous endings at about 150 ms, and reaches its peak amplitude around 250 ms, most clearly over the midline and right posterior sites. The early negativity is followed by an increased negative shift for the anomalous words, with a peak amplitude at approximately 400 ms. This second negativity is larger over posterior than anterior sites and slightly larger over the left than over the right hemisphere. Its topography and latency characteristics are very similar to earlier reports on the N400 effect [17].

Finally, Fig. 1 shows a sustained negativity throughout the whole recording epoch over the anterior sites. This sustained negativity is larger for the anomalous endings. Over posterior sites, the N400 is followed by a late positive shift.

Statistical analyses testing the effects of the sentence-final anomalies were performed for an early and a later window. These analyses were based on the mean amplitudes, relative to a 150 ms pre-target baseline, in a window of 200–300 ms and a window of 350–500 ms after target onset. The mean amplitude values were entered into repeated measures ANOVAs with congruency (correct, anomalous) and site (ATL, Fz, ATR, Cz, PTL, Pz, PTR) as the relevant factors. The Greenhouse-Geisser correction was applied when evaluating

effects with more than one degree of freedom in the numerator [13, 38]. The adjusted degrees of freedom and p-values will be reported. Effects involving the factor site are only reported for significant interactions with congruency.

The ANOVA for the 200–300 ms window resulted in a significant effect of congruency ($F(1,22)=22.01$, $MSe=4.44$, $P < 0.001$). A significant effect of congruency was also obtained in the 350–500 ms window ($F(1,11)=69.97$, $MSe=7.70$, $P < 0.0001$) as well as a significant interaction with site ($F(1,11)=7.60$, $MSe=1.36$, $P < 0.05$).

Fig. 2 shows the topography of the effects in the N250 and the N400 window. This figure shows that the effects in both latency windows are largest over centro-parietal midline sites, and slightly larger over the left than the right hemisphere.

To establish the onset of the effects we performed additional analyses on amplitude measures derived from the subjects' subtraction waveforms. Onset latencies of congruency effects were estimated by first computing the mean amplitude values for the subtraction waveforms in 10 ms bins from target onset until 600 ms after target onset. The values for the resulting 60 latency bins were entered into *t*-tests that tested against the null hypothesis of zero difference from baseline. The onset latency of a given electrode was defined as the bin onset latency at which a *t*-value was obtained at the significance level of 0.05 or better, and which was followed by at least four other bins with significant *t*-values.

The onset latency of the congruency effect at the Cz electrode (where the effect was largest) was 160 ms, and remained significant for the entire testing region (until 600 ms).

3. Experiment 2: sentence-medial semantic anomalies

3.1. Method

Before discussing the results of Experiment 1 we will first present the findings of Experiment 2. In Experiment 2 we tested the effects of the same violations, this time however in sentence-medial positions. A comparison of the results of the two experiments will allow us to answer the question whether the effects observed in Experiment 1 were specific for sentence-final positions. So far, to our knowledge no reports are available in the literature on N400 effects elicited in mid-sentence positions of spoken sentences. A potential problem in recording N400 effects in mid-sentence positions compared to sentence-final positions is the overlap from components elicited by words immediately following the violation in mid-sentence position. Apart from establishing the generalizability of the results of

Experiment 1, we therefore also wanted to establish whether substantial measurement problems occurred in ERP recordings made midway into the speech stream.

The exact same semantic violations as in Experiment 1 were used in this experiment. However, this time the violations occurred in mid-sentence rather than sentence-final position.

3.1.1. Materials

The materials for this experiment were derived from the materials of Experiment 1 by adding sentence continuations of at least three words. To maintain the naturalness of the sentences, for a subset of the original sentences minor additional changes had to be made in the part preceding the critical words. However, the critical words in both the congruent and the anomalous conditions were exactly the same as in Experiment 1. The mean length of the words preceding the critical words was 3.6 phonemes ($SD=2.0$) and 3.0 phonemes ($SD=1.9$) in the congruent and anomalous condition respectively. As in Experiment 1, in the majority of cases these were articles and possessive pronouns.

3.1.2. Subjects

Twelve subjects (eight female; 10 right-handers) participated in this experiment. None of these subjects

had participated in Experiment 1, or in the cloze pretest.

3.2. Results

The overall rejection rate due to trials containing artifacts was 9.2%. Fig. 3 shows the averaged waveforms for the congruent and anomalous mid-sentence targets. Again, as in Experiment 1, N1 and P2 components are very much reduced in the waveform relative to the N1 and P2 elicited by visually presented words.

The waveforms clearly show an early effect of anomaly starting at about 200 ms, with a maximal amplitude at approximately 250 ms for the anomalous condition. This early effect is followed by a large negative shift to the anomalous words, attaining its maximal amplitude at about 400 ms. At frontal electrode sites a sustained negativity to the anomalous items remains visible throughout the whole sentence and recording epoch. In contrast to Experiment 1, at posterior electrode sites there is no late positivity following the N400.

In analyzing the ERP data, the same statistical procedures were used as in Experiment 1. The ANOVA for the early 200–300 ms window resulted in a signifi-

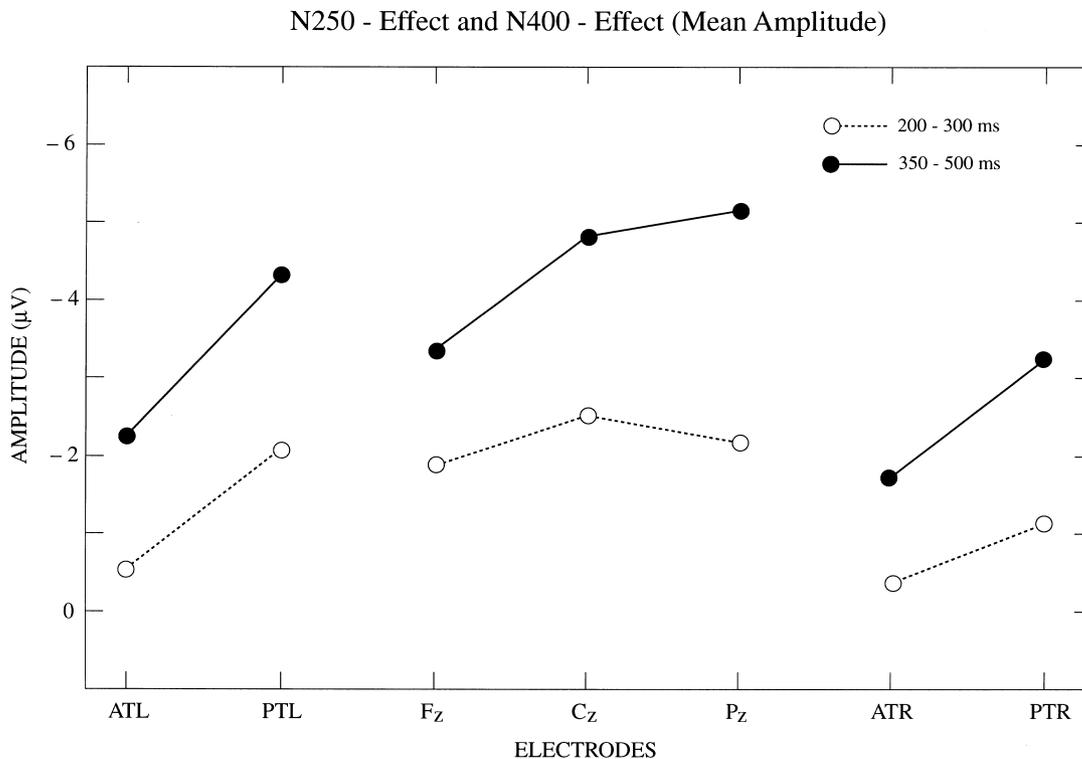


Fig. 2. Connected Speech. The topographical distribution of the N250 and N400 effects for the two left hemisphere sites (ATL, PTL), the three midline sites (Fz, Cz, Pz), and the two right hemisphere sites (ATR, PTR), for the sentence-final semantic anomalies. The N250 effect is determined by subtracting the mean amplitude in the 200–300 ms latency window of the grand average ERP for the semantically anomalous sentence-

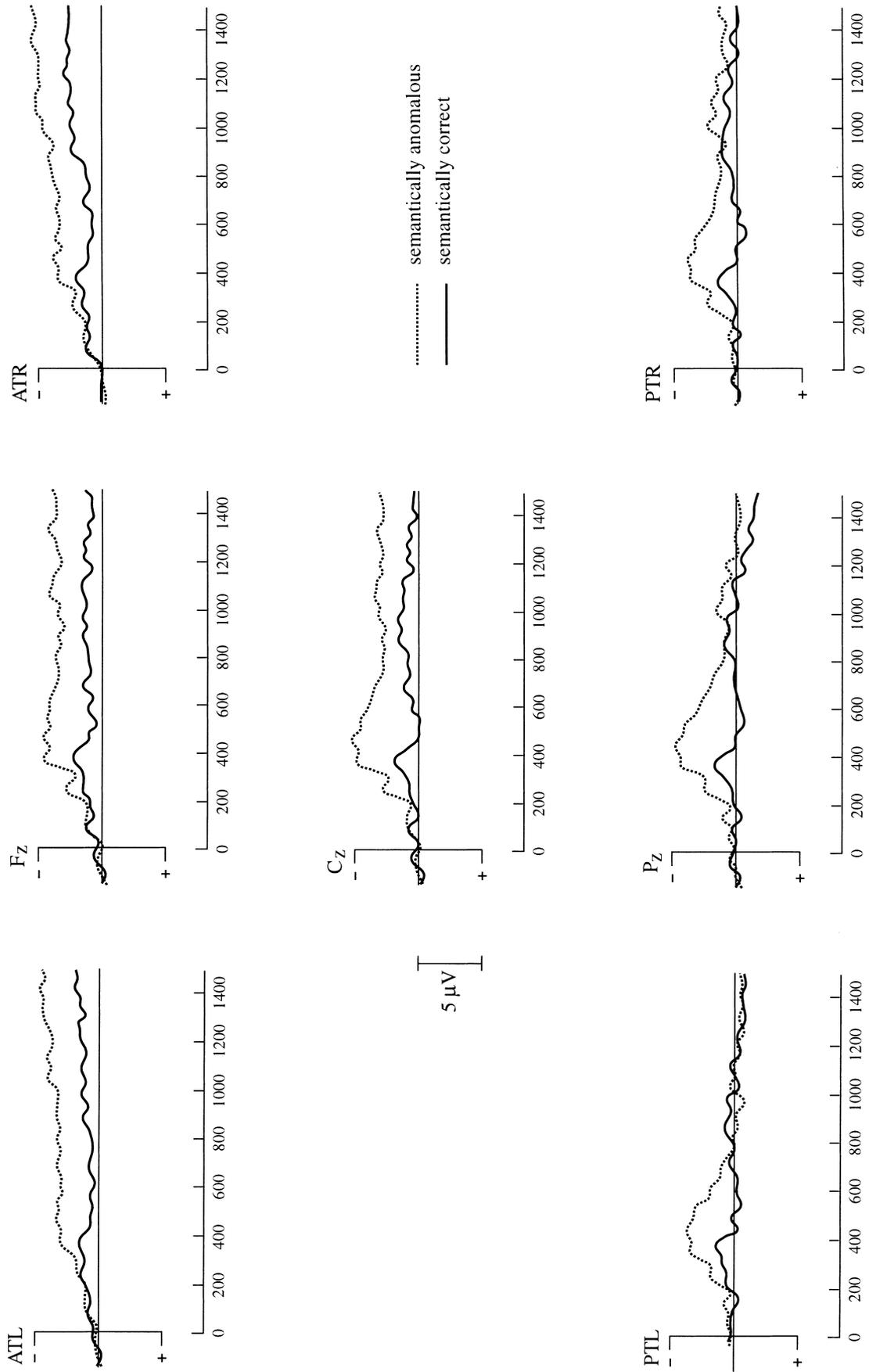


Fig. 3. Connected speech. Grand average ERPs (across all subjects and items), from seven scalp sites, to the mid-sentence semantic anomalies, and to the semantically correct words in mid-sentence position. The onset of the critical words in mid-sentence position is at zero. The time axis is in milliseconds. Each plot represents approximately 654 trials.

cant effect of congruency ($F(1,11)=6.59$, $MSe=6.41$, $P < 0.05$). The ANOVA for the N400 epoch (350–500 ms) also yielded a highly significant congruency effect ($F(1,11)=33.35$, $MSe=10.97$, $P = 0.0001$). No interaction with site was obtained for either of the two windows.

Fig. 4 shows the topography of the congruency effects in the two latency windows.

Again, as in Experiment 1 the effects are slightly larger over posterior than anterior sites, although in this experiment the congruency by site interaction failed to reach significance. In contrast to Experiment 1, this time the effects seemed to be more symmetrically distributed over both hemispheres, with a slight right hemisphere preponderance for the early negativity.

Analyses to establish onset latencies were done in the same way as in Experiment 1. The analyses on the onset latencies and the time course of the congruency effects showed two separate significant regions. At Cz, the early negativity became significant at 240 ms and remained significant until 290 ms. The onset latency of the N400 effect was at 320 ms, with all the remaining bins that were tested (until 600 ms) showing significant t -test results.

4. General discussion

The two experiments on semantic violations in connected speech both resulted in substantial N400 effects time-locked to the word in the sentence that was semantically at odds with the preceding context. This N400 effect was obtained not only to words in sentence-final but also to words in sentence-medial positions. Unlike the prototypical visual N400 effect, the auditory N400 effect was either symmetrical or larger over the left than the right hemisphere. But 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 effect. The hemispheric differences suggest that probably there is also a contribution from non-overlapping neural generators for the two input modalities. However, the distribution of N400 effects observed in reading has varied across experiments, and therefore the distributional deviation of the auditory N400 effects in this experiment from the prototypical visual N400 effect to semantic violations in language, might well fall within the normal range for the visual N400.

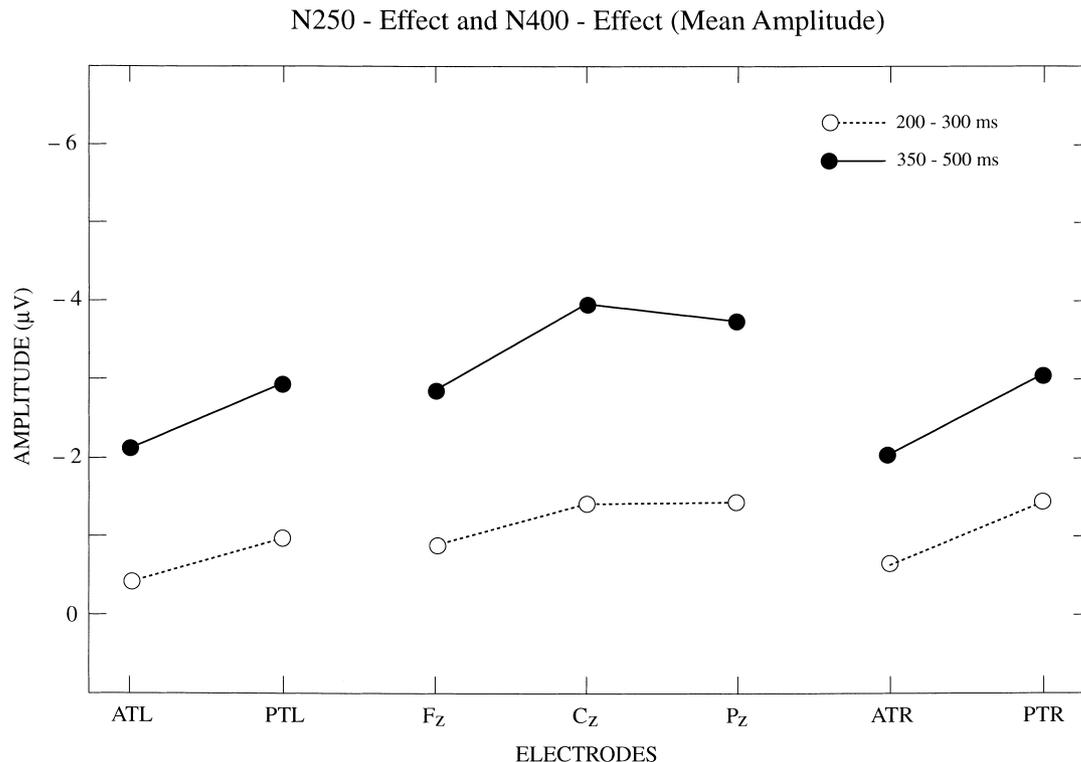


Fig. 4. Connected speech. The topographical distribution of the N250 and N400 effects for the two left hemisphere sites (ATL, PTL), the three midline sites (Fz, Cz, Pz), and the two right hemisphere sites (ATR, PTR), for the mid-sentence semantic anomalies. The N250 effect is determined by subtracting the mean amplitude in the 200–300 ms latency window of the grand average ERP for the semantically anomalous sentence-final words from mean amplitudes of the grand average ERP for their correct counterparts. The N400 effect is determined in the same way on the basis of the mean amplitudes in the 350–500 ms latency window.

Unlike claims in previous studies [17] we did not observe an earlier onset of N400 effects in the auditory compared to the visual modality. Our data rather suggest that in fact the anomalies elicited another effect that preceded the N400 effect in time. The onset of this effect for sentence-final anomalies (150 ms) was quite comparable to what Holcomb and Neville reported for their parietal sites. The lateralization of the earlier effect was not clear-cut. For sentence-final positions it was slightly larger over the left hemisphere, but for mid-sentence positions a small right hemisphere preponderance was observed. Given the relatively small number of electrodes from which recordings were made we are reluctant to attribute functional significance to these subtle differences. Especially the fact that for mid-sentence positions the early and late negativity were clearly separated in time, supports the claim that the N250 and the N400 effects are separate effects.

Although this study was not designed to determine the functional nature of the N250 effect, it clearly is reminiscent of what has been labelled the PMN (Phonological Mismatch Negativity) by Connolly and Phillips [5]. Indirect support for an account in terms of a phonological mismatch between the actual word and the expected lexical candidate comes from the study of Friederici et al. [11]. In this study an N400 effect was present, but the early separation did not occur. A possible explanation for the discrepancy between the Friederici et al. study and the results of our study is that in the former study the semantically anomalous sentence endings all had the same onset as the correct endings, whereas in our case the overwhelming majority (93%) of semantic violations had a different onset cluster than the expected lexical candidate.

In the light of the literature and the present results we put forward the following tentative suggestion about the functional nature of this early negative effect. In spoken word processing information becomes available in a left-to-right order. There is a fair amount of evidence that lexical processing does not wait until the entire speech signal for a particular word is available, but starts immediately on the basis of the initial segments of the spoken word signal [29]. 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, [31]) 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 N250 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 N250 effect is visible relative to a situation in which form-based activation is supported by contextual specifications. This predicts that the N250 effect should vary with the strength of the contextual constraints.

In contrast to the N250 effects, the N400 effects in sentence processing are claimed to arise at the content level only. Once a word is recognized, 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 recognized word against its context that is reflected in the amplitude of the N400. The better the semantic fit, the more reduced the N400 amplitude.

In short, for the processing of spoken words in context, three types of processes have been claimed to be necessary [26, 27, 40]. One is a purely form-based process (lexical access), another is purely content-based (lexical integration). A third, intermediate process is one in which form-based and content-based information are combined to select the appropriate word (lexical selection). We hypothesize that in the context of spoken sentence processing the N250 effect is related to lexical selection and the N400 effect is related to lexical integration.

The interaction between form-based and content-based information is an aspect of both spoken and written sentence processing. Despite this commonality, N250 effects are not standardly observed in visual sentence processing, presumably for at least the following two reasons. First, although there is some evidence that also in the visual modality words are processed from left-to-right [2, 19], in contrast to speech, word information is immediately fully available in writing. Therefore, the separation in time between lexical selection and lexical integration might be substantially smaller in the processing of written sentences than in speech. As a result, possible N250 effects can not be easily separated from N400 effects. Furthermore, to avoid contaminations by eye movements most ERP studies on reading use words that can be read in one fixation. This induces a bias towards relatively short words for the critical comparisons. Longer words might allow for a better temporal separation of lexical selection and integration processes.

Although we acknowledge that our functional account of N250 and N400 effects in the context of the processing of speech remains tentative, we believe that

our data provide strong evidence that auditory ERP effects of semantic violations in sentence context result in an early and a later negative shift, rather than in the monophasic negative shift that has been reported for the visual modality. In a recent study with a 29 electrode montage we have found further evidence for the presence of a negativity preceding the N400, as an indicator of early contextual influences in spoken word recognition [3].

In contrast to other reports in the literature [17, 34], the onset latency of the N400 in speech was very similar to the usual N400 onset latency 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 word recognition point of the critical words in our experiments, the onset of the auditory N400 is actually relatively early.

References

- [1] Ardal S, Donald MW. Long-latency brain responses to incongruent sentences using synthetic speech. *Journal of Psychophysiology* 1991;5:315–26.
- [2] Bergman M, Hudson PTW, Eling PE. How simple complex words can be. *Quarterly Journal of Experimental Psychology* 1988;40:41–72.
- [3] van de Brink D, Brown CM, Hagoort P. The N200 as an electrophysiological indicator of early contextual influences in spoken word recognition. Submitted, 2000.
- [4] Brown C, Hagoort P. The processing nature of the N400: evidence from masked priming. *Journal of Cognitive Neuroscience* 1993;5:34–44.
- [5] Connolly JF, Phillips NA. Event-related potential components reflect phonological, semantic processing of the terminal words of spoken sentences. *Journal of Cognitive Neuroscience* 1994;6:256–66.
- [6] Connolly JF, Stewart SH, Phillips NA. The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain Language* 1990;39:302–18.
- [7] Connolly JF, Phillips NA, Forbes KAK. The effects of phonological semantic features of sentence-ending words on visual event-related potentials. *Electroencephalography, Clinical Neurophysiology* 1995;94:276–87.
- [8] Connolly JF, Phillips NA, Stewart SH, Brake WG. Event-related potential sensitivity to acoustic semantic properties of terminal words in sentences. *Brain Language* 1992;43:1–18.
- [9] Ford JM, Woodward SH, Sullivan EV, Isaacks BG, Tinklenberg JR, Yesavage JA, Roth WT. N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography, Clinical Neurophysiology* 1996;99:235–46.
- [10] Frauenfelder UH, Tyler LK. The process of spoken word recognition: an introduction. *Cognition* 1987;25:1–20.
- [11] Friederici AD, Pfeifer E, Hahne A. Event-related brain potentials during natural speech processing: effects of semantic morphological syntactic violations. *Cognitive Brain Research* 1993;1:183–92.
- [12] Grosjean F. Spoken word recognition processes the gating paradigm. *Perception Psychophysics* 1980;28:267–83.
- [13] Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika* 1959;24:95–112.
- [14] Herning RI, Jones RT, Hunt JS. Speech event related potentials reflect linguistic content processing level. *Brain Language* 1987;30:116–29.
- [15] Holcomb PJ. Semantic priming stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* 1993;30:47–61.
- [16] Holcomb PJ, Neville HJ. Auditory visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Language Cognitive Processes* 1990;5:281–312.
- [17] Holcomb PJ, Neville HJ. Natural speech processing: an analysis using event-related brain potentials. *Psychobiology* 1991;19:286–300.
- [18] Holcomb PJ, Coffey SA, Neville HJ. Visual auditory sentence processing: a developmental analysis using event-related brain potentials. *Developmental Neuropsychology* 1992;8:203–41.
- [19] Hudson PTW, Buijs D. Left-to-right processing of derivational morphology. In: Feldman LB, editor. *Morphological aspects of language processing*. Hillsdale, NJ: Lawrence Erlbaum, 1995. p. 383–96.
- [20] Jasper HH. Report to the committee on methods of clinical examination in electroencephalography. Appendix: the ten-twenty system of the International Federation. *Electroencephalography, Clinical Neurophysiology* 1958;10:371–5.
- [21] Kellenbach ML, Michie PT. Modulation of event-related potentials by semantic priming: effects of color-cued selective attention. *Journal of Cognitive Neuroscience* 1996;8:155–73.
- [22] Kutas M, Hillyard SA. Reading senseless sentences: brain potentials reflect semantic anomaly. *Science* 1980;207:203–5.
- [23] Kutas M, Van Petten C. Electrophysiological perspectives on comprehending written language. In: Rossini PM, Maguère F, editors. *New trends, advanced techniques in clinical neurophysiology (EEG Suppl 41)*. Amsterdam: Elsevier, 1990. p. 155–67.
- [24] Kutas M, Van Petten CK. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher MA, editor. *Handbook of psycholinguistics*. San Diego: Academic Press, 1994. p. 83–143.
- [25] Marslen-Wilson WD. Function, process in spoken word-recognition. In: Bouma H, Bouwhuis DG, editors. *Attention, performance X: control of language processes*. Hillsdale, NJ: Erlbaum, 1984. p. 125–50.
- [26] Marslen-Wilson WD. Functional parallelism in spoken word-recognition. *Cognition* 1987;25:71–102.
- [27] Marslen-Wilson WD. Access, integration: projecting sound onto meaning. In: Marslen-Wilson WD, editor. *Lexical representation, process*. Cambridge, MA: MIT Press, 1989. p. 3–24.
- [28] Marslen-Wilson WD, Tyler LK. The temporal structure of spoken language understanding. *Cognition* 1980;8:1–71.
- [29] Marslen-Wilson WD, Welsh A. Processing interactions, lexical access during word-recognition in continuous speech. *Cognitive Psychology* 1978;10:29–63.
- [30] McCallum WC, Farmer SF, Pocock PV. The effects of physical, semantic incongruities on auditory event-related potentials. *Electroencephalography, Clinical Neurophysiology* 1984;59:477–88.
- [31] Norris D. Shortlist: a connectionist model of continuous speech recognition. *Cognition* 1994;52:189–234.
- [32] O'Halloran JP, Isenhardt R, Sandman CA, Larkey LS. Brain responses to semantic anomaly in natural continuous speech. *International Journal of Psychophysiology* 1988;6:243–54.

- [33] Oldfield RC. The assessment, analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113.
- [34] Osterhout L, Holcomb PJ. Event-related potentials elicited by syntactic anomaly. *Journal of Memory, Language* 1992;31:785–806.
- [35] Osterhout L, Holcomb PJ. Event-related potentials, language comprehension. In: Rugg MD, Coles MGH, editors. *Electrophysiology of mind: event-related brain potentials, cognition*. Oxford: Oxford University Press, 1995.
- [36] Swaab T, Brown C, Hagoort P. Spoken sentence comprehension in aphasia: event-related potential evidence for a lexical integration deficit. *Journal of Cognitive Neuroscience* 1997;9:39–66.
- [37] Van Petten C, Coulson S, Rubin S, Plante E, Parks M. Time course of word identification, semantic integration in spoken language. *Journal of Experimental Psychology: Learning Memory, Cognition* 1999;25:394–417.
- [38] Winer BJ. *Statistical principles in experimental design*. 2nd ed. New York: McGraw-Hill, 1971.
- [39] Woodward SH, Ford JM, Hammett SC. N4 to spoken sentences in young, older subjects. *Electroencephalography, Clinical Neurophysiology* 1993;87:306–20.
- [40] Zwitserlood P. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 1989;32:25–64.