Seeing and Hearing Meaning: ERP and fMRI Evidence of Word versus Picture Integration into a Sentence Context

Roel M. Willems[1], Aslı Özyürek[1,2], and Peter Hagoort[1,2]

Abstract

■ Understanding language always occurs within a situational context and, therefore, often implies combining streams of information from different domains and modalities. One such combination is that of spoken language and visual information, which are perceived together in a variety of ways during everyday communication. Here we investigate whether and how words and pictures differ in terms of their neural correlates when they are integrated into a previously built-up sentence context. This is assessed in two experiments looking at the time course (measuring event-related potentials, ERPs) and the locus (using functional magnetic resonance imaging, fMRI) of this integration process. We manipulated the ease of semantic integration of word and/or picture to a previous sentence context to increase the semantic load of processing. In the ERP study, an increased semantic load led to an N400 effect which was similar for pictures and words in terms of latency and amplitude. In the fMRI study, we found overlapping activations to both picture and word integration in the left inferior frontal cortex. Specific activations for the integration of a word were observed in the left superior temporal cortex. We conclude that despite obvious differences in representational format, semantic information coming from pictures and words is integrated into a sentence context in similar ways in the brain. This study adds to the growing insight that the language system incorporates (semantic) information coming from linguistic and extralinguistic domains with the same neural time course and by recruitment of overlapping brain areas. ■

INTRODUCTION

Understanding language always occurs within a situational context, such as knowledge about the person you are talking to or the location one is in (Clark, 1996). Therefore, understanding language often implies combining streams of information from different modalities. For instance, consider a biology teacher describing the properties of an animal while at the same time showing a slide with a picture of the animal. In such a case, auditory and visual information do not bear a direct physical connection. That is, the string of sounds describing a concept never directly maps onto the visual appearance of this concept. This raises the question how verbal (linguistic) and visual (extralinguistic) semantic information combine to form a coherent interpretation of a speaker's message in relation to the overall context. The current study investigates this question by assessing the neural integration of semantic information from words and pictures embedded within a spoken sentence context. Our main aim was to investigate whether, despite differences in representational format, semantic information from pictures and words is integrated into an overall representation of an utterance in the same way as unimodal semantic information from a word is. Within the broader context of language comprehension, we wanted to investigate differences and commonalities between linguistic and extralinguistic information processing during sentence comprehension.

We investigated both the neural time course (measuring event-related potentials, ERPs) and the locus (using functional magnetic resonance imaging, fMRI) of this multimodal integration process. Combining temporal and spatial neural information in this way provides a more complete view of the integration process under study.

Lexical Semantic Information in a Sentence Context

The seminal ERP studies by Kutas and Hillyard (1980, 1984) showed that words that are semantically anomalous in relation to the preceding sentence context lead to a more negative deflection in the ERP waveform than words that are semantically congruent. For example, the sentence-final word of the sentence, "She spread her bread with *socks*," leads to a negative deflection in the ERP waveform in comparison to the ERP for a congruous ending as in "She spread her bread with *butter*." This N400 effect occurs between 250 and 500 msec after the anomalous word, and is usually maximal at central–posterior electrode sites. N400 effects are also observed

---

[1]Radboud University Nijmegen, Nijmegen, The Netherlands, [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

when a critical word is a possible but unexpected continuation of a sentence, without being a semantic violation (Hagoort & Brown, 1994; Kutas & Hillyard, 1984). The N400 has become an established ERP component which is thought to reflect the ease of integration of a word into a preceding context (see Brown, Hagoort, & Kutas, 2000; Kutas & Van Petten, 1994 for reviews).

fMRI studies of sentences with semantic anomalies comparable to those used in N400 studies have reported increased activation in left inferior frontal and/or left temporal areas (Willems, Özyürek, & Hagoort, 2007; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Kuperberg et al., 2000, 2003; Baumgaertner, Weiller, & Buchel, 2002; Ni et al., 2000). Based on these findings, it has been claimed that these areas are involved in semantic integration because they respond to a higher load of integration elicited by the difficulty of semantic processing. Recent work has indeed shown that increased semantic ambiguity without semantic anomalies also leads to increased activations in left inferior frontal and left temporal regions (Rodd, Davis, & Johnsrude, 2005; Davis et al., 2007).

Extralinguistic Information in a Sentence Context

The integration of extralinguistic information into a preceding context has been explored in a variety of ways in ERP or fMRI studies (e.g., Hagoort et al., 2004; Koelsch et al., 2004; Van Petten & Rheinfelder, 1995). A complete review of how extralinguistic information influences sentence comprehension is beyond the scope of this article. We therefore restrict our focus on studies investigating the integration of visual information into a preceding (sentence) context.

Few studies have looked into the integration of picture information into a sentence context. In an ERP study, Ganis, Kutas, and Sereno (1996) presented sentences that either ended with a word or a picture that could be anomalous or not. Similar N400 effects were found to anomalous words and pictures. However, the scalp distribution for the anomalous pictures was more frontal than for the anomalous words. Nigam, Hoffman, and Simons (1992) also found similar N400 effects for pictures and words, but did not find a difference in scalp distribution. However, this might be due to the limited number of electrodes that they recorded from, which did not cover the frontal part of the brain. Federmeier and Kutas (2001) found a correlation between the amplitude of the N400 effect and the semantic fit of a picture with respect to the preceding part of a sentence. Again, there was a frontal scalp distribution for the effects. Additionally, they observed an N300 effect to the anomalous pictures. Some other ERP studies have investigated the processing of visual information following a *visual* context instead of a language context. West and Holcomb (2002), for instance, presented a series of pic-

tures forming a simple story. The last picture was either a congruous or an incongruous ending of the story. Incongruous pictures elicited increased N300 and N400 effects, with a maximal distribution over centro-frontal electrodes. Sitnikova, Kuperberg, and Holcomb (2003) had congruous or incongruous objects appear in video clips of real-world events. They observed an N400 effect for the incongruous objects with a fronto-central maximum in the scalp distribution. Finally, Ganis and Kutas (2003) had congruent or incongruent objects appear in still images of real-world events. An increased negativity strongly resembling the N400 was observed for the incongruous as compared to the congruous objects.

Several priming studies report similar findings when pictures and words are presented outside of a sentence context. That is, N300 and N400 effects are reported to incongruous picture pairs, with a more frontal scalp distribution than is normally seen for word–word priming studies (McPherson & Holcomb, 1999; Holcomb & McPherson, 1994; Barrett & Rugg, 1990).

In summary, ERP studies manipulating the semantic fit of pictures in relation to a (sentence) context report similar N400 amplitudes and onset latencies as found for integration of semantic information conveyed through a word. Differences are reported, however, in scalp distribution which is more frontal for pictures than for words, and in the finding of an earlier separate negativity, the N300. The latter component has been suggested to reflect the degree of effort needed to integrate an object-specific/imagistic representation into a preceding context (e.g., McPherson & Holcomb, 1999).

From neuroimaging studies, little is known about the neural localization of sentence-level processing of visual extralinguistic information. In an earlier study, we looked at how meaningful co-speech gestures compare to spoken words when anomalous within a sentence context (Willems et al., 2007). Overlap between lexical violations and gesture violations was found in the left inferior frontal cortex. Relevant to the present study is the considerable literature on the neural correlates of the semantic representation of visually presented objects. Such studies mostly find that the ventral temporal cortex is activated to the perception of a large variety of objects (Martin & Chao, 2001; Schacter & Buckner, 1998). More important for the present study is that many of these studies also report the inferior frontal cortex to be sensitive to the repeated presentation of an object (Martin & Chao, 2001; Schacter & Buckner, 1998) or of a word and an object (e.g., Lebreton, Desgranges, Landeau, Baron, & Eustache, 2001). A commonly held view derived from these and other studies is that ventral temporal activation is related to semantic knowledge of an object, whereas inferior frontal activation is related to processes of semantic selection or retrieval (Martin & Chao, 2001; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner, Desmond, Demb, Glover, & Gabrieli, 1997). We

would like to point out that although related to our study, the priming studies of objects differ in important aspects from the present study. In the present paradigm, there is a relatively rich linguistic context to which picture or word can be integrated. In the repeated presentation of exemplars of object categories, this is arguably not the case.

To our knowledge, the present study is the first to investigate the integration of pictures at the sentence level, using both fMRI and ERPs.

The Present Study

Within the study of the cognition of language, the issue of how linguistic and extralinguistic information are integrated into a sentence context is reflected in the distinction between one-step and two-step models of language comprehension. The implication of two-step models is that, first, the meaning of a sentence is computed and second, the sentence meaning is integrated with extralinguistic information such as information about the speaker's identity (e.g., Lattner & Friederici, 2003; Cutler & Clifton, 1999). This position is a consequence of Fregean compositionality, which states that the meaning of an utterance is a function of the meaning of its parts and of the syntactic rules by which these parts are combined (see Culicover & Jackendoff, 2006). Because the domain of syntactic rules is the sentence, the implication of this idea is that language interpretation takes place in a two-step fashion. It is important for the present study that the two-step model at least implies that linguistic computation should precede the integration of nonlinguistic information in time (see Hagoort & van Berkum, 2007 for further discussion). Adherents of a one-step model, in contrast, take as their starting point the "immediacy assumption," that is, the idea that *every* source of information that constrains the interpretation of an utterance (syntax, prosody, word-level semantics, prior discourse, world knowledge, knowledge about the speaker, gestures, etc.) can, in principle, do so immediately (Hagoort & van Berkum, 2007; Spivey Knowlton & Sedivy, 1995; Tanenhaus, Spivey Knowlton, Eberhard, & Sedivy, 1995; Tanenhaus & Trueswell, 1995; Trueswell & Tanenhaus, 1994; Taraban & McClelland, 1990). Summarized, proponents of a two-step model would expect indicators of semantic integration in the ERP to be manifested earlier when a word has to be integrated as compared to when a picture has to be integrated into the previous sentence context.

In terms of cortical areas important for language comprehension, a recent neurobiological account of language comprehension has argued for the left inferior frontal cortex to be a general (i.e., not domain-specific) unification site (Hagoort, 2005a, 2005b). Unification entails integration of information into a built-up representation of the previous sentence context as well as a selection of appropriate candidates for integration (Hagoort, 2005a, 2005b). When unification is more difficult, more resources are needed to integrate linguistic as well as extralinguistic information, resulting in increased activation levels in the left inferior frontal cortex. If this is indeed the case, we should observe increased activation both when a picture and when a word are harder to integrate. However, if this area's role is restricted to integrating language information, no such increase should be observed when information conveyed through a picture has to be integrated.

On the basis of previous studies investigating sentence-level integration of co-speech gestures compared to words, we have argued for linguistic and extralinguistic information to be integrated in the same way into a linguistic context (Özyürek, Willems, Kita, & Hagoort, 2007; Willems et al., 2007). In these studies, the semantic fit of a word or of a co-occurring co-speech gesture to the preceding sentence context was manipulated. Mismatching spoken words and co-speech gestures elicited N400 effects with similar onset latencies. Although co-speech gestures and the pictures that we investigate here are both extralinguistic information, clear differences exist as well. One important characteristic of co-speech gestures is that their meaning is not recognized unambiguously when presented outside of a language context (Krauss, Morrel Samuels, & Colasante, 1991). In contrast, pictures can stand on their own. Therefore, the present study is a stronger test for the claim that semantic integration at the sentence level is not domain (i.e., language) specific. If neural correlates of integration of pictures and words are similar, it follows that also information that is not necessarily bound to a language context is integrated with the same spatio-temporal profile in the brain as linguistic information is.

To address these questions, we presented participants with spoken sentences in which a critical word was manipulated to either fit the sentence context or not. The critical words were accompanied by pictures (i.e., line drawings) that could also either match or mismatch with regard to the previous part of the sentence. This manipulation resulted in four conditions (see Table 1): correct condition [Picture (P) +, Language (L) +]; language mismatch condition (P+L−); picture mismatch condition (P−L+); double mismatch condition (P−L−). In the language mismatch, the critical word was harder to integrate semantically into the preceding sentence context, whereas the co-occurring picture matched the sentence context. In the picture mismatch condition, the picture was harder to integrate into the previous context, whereas the critical word matched the spoken sentence context. In the double mismatch condition, both the picture and the word were difficult to integrate into the previous sentence context. Note that in the language and picture mismatch conditions the critical word and the overlapping picture locally mismatched (e.g., picture CHERRY, word "flower," and vice-versa), whereas in the double mismatch condition they locally matched (e.g., both "cherry"). This manipulation

Table 1. An Example of the Stimulus Materials

Dutch: De man gaf zijn vrouw een mooie bloem/kers die avond

English: "The man gave his wife a nice flower/cherry that evening"

Correct condition

P+L+: The man gave his wife a nice flower that evening

Language mismatch

P+**L**−: The man gave his wife a nice **cherry** that evening

Picture mismatch

P−L+: The man gave his wife a nice flower that evening

Double mismatch

P−**L**−: The man gave his wife a nice **cherry** that evening

Pictures were displayed time-locked to the onset of the noun (underlined). Note that the condition coding (P+L+, P+L−, etc.) refers to the match/mismatch of either the noun (Language: L) or the Picture (Picture: P) to the part of the sentence preceding the word that is underlined, with a minus sign indicating a mismatch. That is, in the correct condition (P+L+), both the word "flower" as well as the picture [FLOWER] fit the preceding sentence context. In the language mismatch condition (P+L−), the word "cherry" does fit the preceding sentence context less well, whereas the picture [FLOWER] does fit. Conversely, in the picture mismatch condition (P−L+), the picture [CHERRY] does not fit the preceding sentence context, whereas the word "flower" does fit. Finally, in the double mismatch condition (P−L−), both the word "cherry" and the picture [CHERRY] do not fit the preceding sentence context. Mismatching words are indicated in **bold**. All stimuli were in Dutch.

enabled us to distinguish integration at the "local" level of simultaneously occurring word and picture from integration at the "global" sentence level; that is, integration into a higher-level representation built-up on the basis of the preceding context information.

We had three specific hypotheses. First, for the ERP data we hypothesized that manipulating the match of both picture and word would lead to an N400 effect comparable in size and onset latency. Moreover, we were curious to see if an N300 effect would be apparent and if so, whether it would be specific to the picture mismatch condition. Previous ERP studies have mostly

compared the presence of an N300 effect in reaction to pictures to the absence of an N300 to words indirectly. That is, in most studies, either words or pictures were presented. Our design allows for assessing the functional relevance of the N300 in the sense that if it is sensitive to semantic load of a picture it should occur in the picture mismatch condition but not in the language mismatch condition. Second, for the fMRI study, we predict a stronger involvement of the inferior frontal cortex in both the picture and word mismatch conditions. If so, this would be evidence that this region, besides its well-established role in spoken and written language comprehension (e.g., Vigneau et al., 2006; Bookheimer, 2002), also takes extralinguistic visual information into account during language comprehension. Third, on the basis of earlier findings (Özyürek et al., 2007), we hypothesized our findings to reflect semantic processing at the "global" sentence level but not at the "local" level of the simultaneous picture and word. In relation to this, it should be noted that recent data acquired in our lab show that without an elaborate context, simultaneously presented incongruent word and object pairs do not elicit an N400 (Özyürek, Willems, & Hagoort, unpublished results).

Overall, our main question regards the similarity or dissimilarity of integrating linguistic and extralinguistic information into a sentence context. Differences in neural indicators of semantic processing would favor an account in which linguistic information has a preferred status in sentence integration (Fodor, 1983; Forster, 1979), whereas findings of similar neural correlates would support the idea that linguistic and extralinguistic information are integrated with a similar neural time course and by recruiting overlapping cortical areas (Hagoort & van Berkum, 2007). Moreover, an earlier effect to words than to pictures would be in favor of two-step models of language comprehension (e.g., Lattner & Friederici, 2003; Cutler & Clifton, 1999), whereas similar neural time courses would favor accounts of immediacy in which a broad range of information types is immediately incorporated into a discourse model (Hagoort & van Berkum, 2007; Spivey Knowlton & Sedivy, 1995; Tanenhaus & Trueswell, 1995; Tanenhaus et al., 1995; Trueswell & Tanenhaus, 1994; Taraban & McClelland, 1990).

METHODS

Materials and Procedure

A total of 328 sentences (mean duration 3196 msec, range 2164–4184 msec) were recorded in a sound-attenuated room at 44.1 kHz, spoken at a normal rate by a native Dutch female speaker. Half of these sentences differed in one critical word, which was never in sentence-final position. In each sentence, a short context was introduced to which the critical word could fit more or less easily. Critical words were nouns that corresponded to the names given by a separate group of participants ($n =$

32) to a large set of black and white line drawings. All critical words had a picture equivalent with a naming consistency of 85% or higher. In total there were 26 critical words with their picture equivalents. All words were one syllable long and started with a plosive consonant. Every critical word occurred equally often in a matching and a mismatching sentence context. The critical word in the mismatching sentence always had a different onset consonant than the critical word in the semantically correct sentence. Sentences were pretested in a cloze probability test that was given to a separate participant group ($n = 16$). The percentage of participants that gave the target word as response was taken as a measure of its cloze probability. Overall, the mean cloze probability was 16% for the matching critical words (range 0–69%), and 0% for the semantically anomalous critical words. We choose for critical words with low cloze probabilities to avoid confounding effects of prediction (e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Our manipulation resulted in four conditions (Table 1): (i) correct condition (P+L+); (ii) language mismatch (P+L−); (iii) picture mismatch (P−L+); (iv) double mismatch (P−L−). Note that mismatch in these materials is always defined relative to the preceding sentence context.

Four stimulus lists of 164 trials each were created in which only one item of every stimulus quartet (as in Table 1) was presented. Sentences were pseudorandomized with the constraint that the same condition occurred maximally two times in a row. Every list contained an equal amount of stimuli from the four conditions (41 per condition). Every target word and picture was repeated, on average, 6.3 times (range = 5–8, modus = 6, median = 6 repetitions) in every stimulus list. Pictures were presented from the onset of the critical word to the end of the sentence.

EXPERIMENT 1: EEG

Participants

Twenty-four healthy right-handed (Oldfield, 1971) participants with Dutch as their mother tongue took part in the electroencephalogram (EEG) study. None had any known neurological history or hearing complaints, and all had normal or corrected-to-normal vision. Eight participants' data had to be discarded because of an excessive number of blinks and eye movements, leaving datasets from 16 participants (mean age = 22.4 years, range = 18–34, 11 women). Subjects were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the Declaration of Helsinki.

Procedure

Stimuli were presented using Presentation software (version 9.13, www.neuro-bs.com/). Pictures had varying sizes depending upon the object they represented

and were maximally 8 × 8 cm, shown at a viewing distance of 90 cm (5° × 5° visual angle). A trial started with 600 msec blank screen, followed by the spoken sentence and the picture, 1000 msec blank screen, and 2500 msec with a fixation cross on the screen. Participants were instructed to sit still in a comfortable position and to blink only when a fixation cross was presented. The test session started with eight trials which contained different critical words than used in the main part of the experiment. Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterward. At the end of the test session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

Recording and Analysis

The EEG was recorded from 27 electrode sites across the scalp using an Electrocap with Ag/AgCl electrodes, each referred to the left mastoid. Electrodes were placed on standard electrode sites (Fz, FCz, Cz, Pz, Fp2, F3, F4, F8, F7, FC5, FC1, FC2, FC6, T7, T8, C3, C4, CP5, CP1, CP2, CP6, P7, P3, P4, P8, O1, O2). Vertical eye movements and blinks were monitored by means of two electrodes, one placed beneath and one above the left eye. Horizontal eye movements were monitored by means of a left to right bicanthal montage. Activity over the right mastoid was recorded to determine whether there were additional contributions of the experimental variables to the two presumably neutral mastoid sites. No such differences were observed. The EEG and electrooculogram (EOG) recordings were amplified with BrainAmp DC amplifiers, using a band-pass filter from 10 sec to 100 Hz. Impedances were kept below 5 kΩ for all channels. The EEG and EOG signals were recorded and digitized using Brain Vision Recorder software (version 1.03), with a sampling frequency of 500 Hz.

The data were filtered off-line with a 30-Hz low-pass filter, re-referenced to the mean of the two mastoids, and segmented from 150 msec before to 1000 msec after the critical word. Segments were normalized to the mean amplitude of a baseline period 150 msec before the critical word (baseline correction). All segments were screened for eye movements, electrode drifting, amplifier blocking, and muscle artifacts. Trials containing such artifacts were rejected (mean = 8.6%, SD = 5.2%, range = 0–18%). Rejected trials were equally distributed across conditions ($F < 1$). Segments were averaged for each condition for each participant at each electrode site. Repeated measures analysis of variance (ANOVA) was applied to the mean activity in four time windows (see Results) with factors condition (P+L+, P+L−, P−L+, P−L−) and quadrant (left anterior, right anterior, left posterior, right posterior). Electrodes were assigned to quadrants as follows: left anterior (F3, F7, FC1, FC5, C3), right anterior (F4, F8, FC2, FC6, C4), left posterior

(CP1, CP5, P3, P7, O1), and right posterior (CP2, CP6, P4, P8, O2). A separate ANOVA was performed for the midline electrodes (Fz, FCz, Cz, Pz). Huynh–Feldt correction for violation of sphericity assumption was applied when appropriate (Luck, 2005; Picton et al., 2000; Huynh & Feldt, 1976). Differences in N400 effect onset latencies were tested by calculating the time bin (bins of 10 msec) in which 20% of the total area of the difference waves of the experimental conditions with the correct condition in the 200–500 msec latency window was reached (fractional area latency analysis). Statistical significance of these differences was assessed by using the jackknifing procedure described by Miller, Patterson, and Ulrich (1998).

Results

The grand-average waveforms (Figure 1) show clear N1 and P2 components followed by a negativity starting from 350 msec resembling the N400. The correct condition showed a slightly decreased peak at the N1 and a stronger positivity at the P2 peak compared to the other three conditions (Figures 1 and 2A). In the N400 time window, the three mismatch conditions (P+L−, P−L+, P−L−) were more negative than the correct condition. The mismatch conditions stayed more negative than the correct condition until the end of the segment. Consequently, ANOVAs were done on the mean amplitudes in the latency ranges 170–250 msec, 300–550 msec, and 600–900 msec. Additional tests were performed in the N300 time window (225–325 msec) given that previous literature (e.g., McPherson & Holcomb, 1999) reports picture-specific effects in this time window.

P2 Time Window (170–250 msec)

Statistical analyses in this time window failed to reveal a significant effect of condition [$F(3, 45) = 2.34, p = .09$]. There was also no Condition × Quadrant interaction ($F < 1$). However, in the ANOVA over midline electrodes, a main effect of condition was found [$F(3, 45) = 2.87, MSE = 9.97, p = .047$]. Planned comparisons over the midline electrodes, of every experimental condition versus the correct condition, showed this effect to be strongest in the picture mismatch condition [$F(1, 15) = 5.82, MSE = 26.96, p = .029$], although there were also marginally significant effects in the language mismatch versus correct condition [$F(1, 15) = 4.29, MSE = 21.27, p = .056$] and in the double mismatch versus correct condition comparisons [$F(1, 15) = 4.51, MSE = 10.75, p = .051$].

N300 Time Window (225–325 msec)

The morphology of the grand-average waveforms does not clearly indicate the presence of a separate N300 component. Given previous findings of the N300 for mis-

matching pictures, we tested effects in the 225–325 msec time window (e.g., McPherson & Holcomb, 1999). A main effect of condition was observed [$F(3, 45) = 3.17, MSE = 42.96, p = .040$], but no Condition × Quadrant interaction [$F(9, 135) = 1.26, MSE = 10.22, p = .293$]. Pairwise comparisons between all conditions were performed and p values were corrected for the number of tests accordingly. Only the double mismatch versus correct condition differed significantly from each other [$F(1, 15) = 9.82, MSE = 17.3, p = .042$]. Since previous studies found the distribution of the N300 effect to be frontal, we separately tested in left and right anterior quadrants. Again, there was a main effect of condition [left: $F(3, 45) = 3.47, MSE = 22.54, p = .040$; right: $F(3, 45) = 4.22, MSE = 11.88, p = .010$]. Pairwise comparisons revealed only the double mismatch condition to be significantly different from the correct condition in the left anterior quadrant [$F(1, 15) = 11.33, MSE = 10.17, p = .024$]. The picture mismatch versus correct condition was marginally significant in the left anterior quadrant only [$F(1, 15) = 8.20, MSE = 34.96, p = .073$]. No other comparisons revealed significant differences between conditions. Summarized, although there is a main effect of condition in this time window, this effect is not specific to the picture and/or double mismatch conditions, as might have been expected based upon previous findings.

N400 Time Window (300–550 msec)

Table 2 summarizes the results in this time window. There was a main effect of condition [$F(3, 45) = 11.46, p < .001$], but no Condition × Quadrant interaction [$F(9, 135) = 1.56, p = .16$]. To explore specific differences between conditions, pairwise comparisons were performed. Accordingly, the p values are corrected for the number of tests performed (see Table 2). Pairwise comparisons revealed that all mismatch conditions differed significantly from the correct condition. No other comparisons showed significant differences between conditions (Table 2). These effects are spatially smeared out over the scalp, with a tendency for fronto-central electrodes to show the greatest effect size (Figure 2B). To formally test the onset latencies of the N400 effects, a fractional area latency measure was computed in the 200–500 msec time window (see EEG Methods). The time point at which 20% of the grand-average difference waveform was reached was 305 msec for the language mismatch–correct condition, 285 msec for the picture mismatch–correct condition, and 295 msec for the double mismatch–correct condition. Statistical testing of these differences (see Miller et al., 1998) revealed that onset latencies did not differ from each other (all $t < 1$).

Finally, the scalp distributions of the difference waves (Figure 2B) were tested in a two-way ANOVA with factors difference (language mismatch–correct condition, picture mismatch–correct condition, double mismatch–correct condition) and quadrant. The lack of an interaction effect
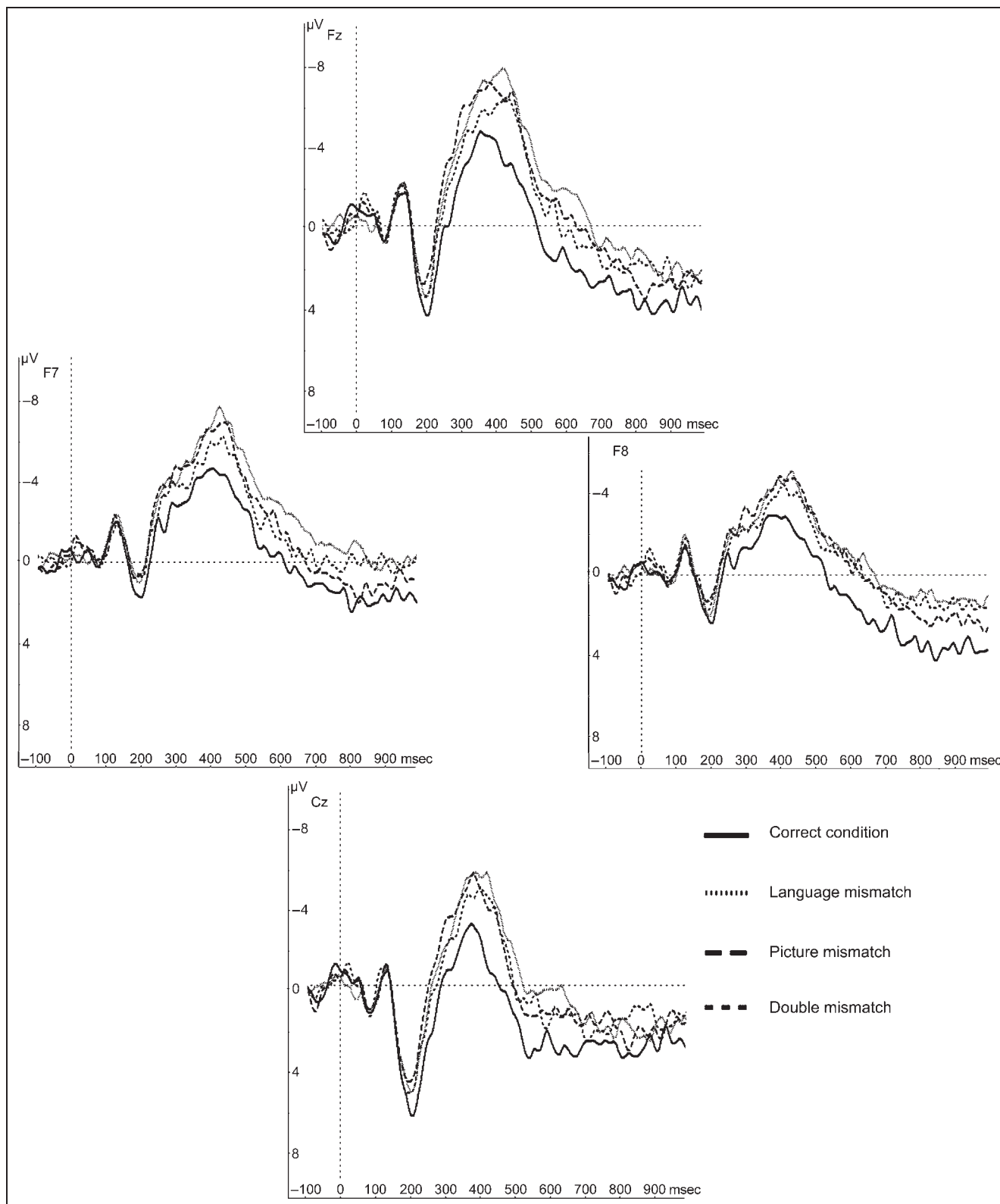
Figure 1. Grand-average ERPs for the four conditions at electrodes Fz, F7, F8, and Cz. ERPs were time-locked to the onset of the critical word and picture. Negativity is plotted upward.

(F < 1) indicates that the scalp distributions were similar for the N400 effects elicited by the mismatch conditions.

Late Time Window (600–900 msec)

In the late time window (600–900 msec), there was a significant main effect of condition [$F(3, 45) = 3.71$, $MSE = 31.55$, $p = .018$] and a significant Condition × Quadrant interaction [$F(9, 135) = 3.62$, $MSE = 6.26$, $p = .003$]. Subsequent tests in specific quadrants revealed main effects of condition only in the left anterior [$F(3, 45) = 5.82$, $MSE = 13.91$, $p = .002$] and right anterior quadrants [$F(3, 45) = 5.80$, $MSE = 11.97$, $p = .002$]. Again, we tested pairwise comparisons between all
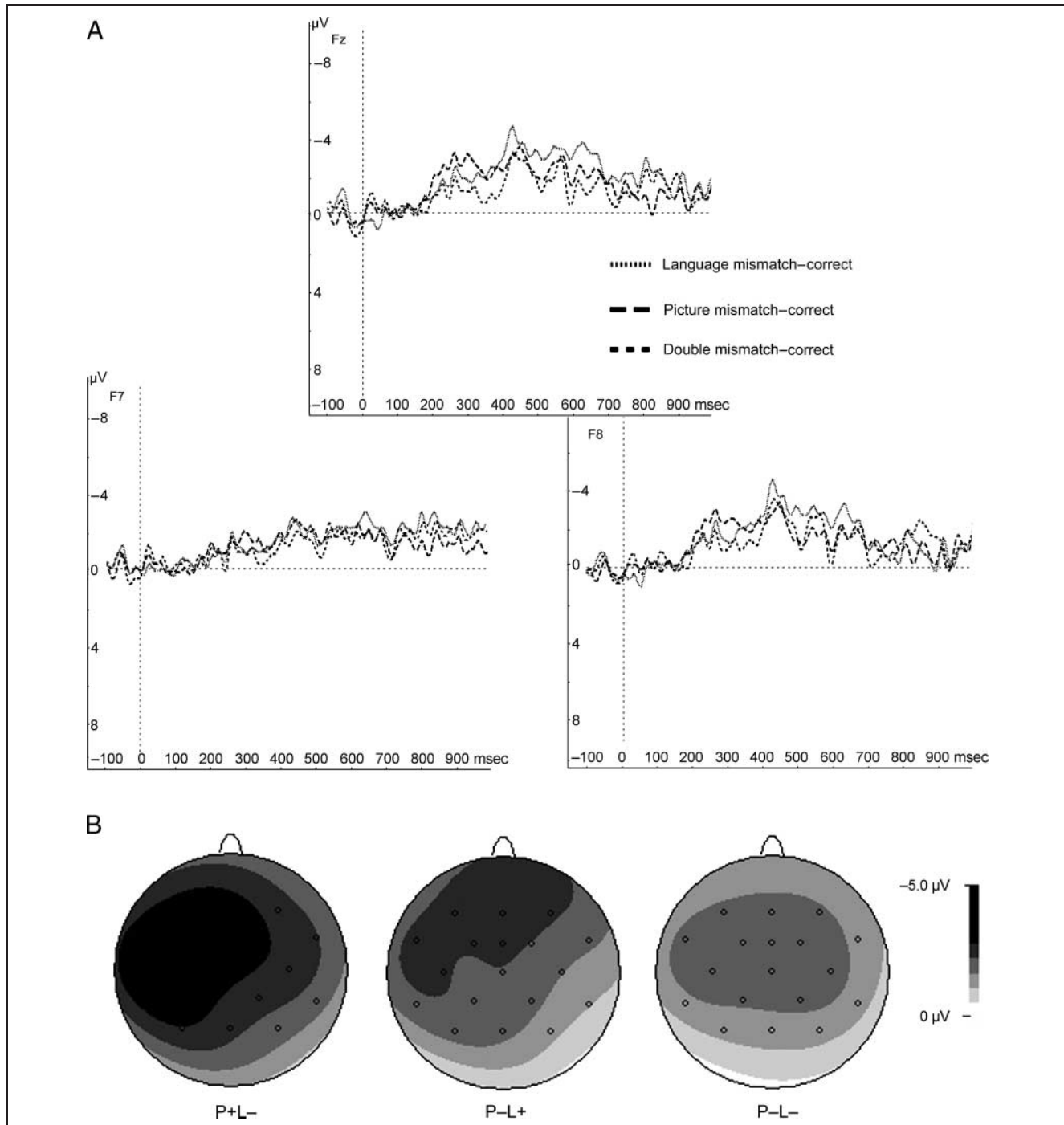
Figure 2. (A) Difference waves of the experimental conditions minus the correct condition (language mismatch–correct condition; picture mismatch–correct condition; double mismatch–correct condition) at electrodes Fz, F7, and F8. Difference waves are time-locked to the onset of the critical word and picture. Negativity is plotted upward. (B) Spline interpolated isovoltage maps of the mean difference wave 300–550 msec after the critical word. Displayed are the difference of the language mismatch condition (left), picture mismatch condition (middle), and double mismatch condition (right) with the correct condition.

Table 2. ERP Results in the 300–550 msec Time Window

Source	df	F	MSE	p
ANOVA (4 Conditions × 4 Quadrants)				
Condition	3, 45	11.46	26.72	< .001***
Condition × Quadrant	9, 135	1.56	8.86	.162
Pairwise comparisons				*p*(corr.)
P+L− vs. P+L+	1, 15	42.34	39.18	< .001***
P−L+ vs. P+L+	1, 15	11.42	73.08	.025*
P−L− vs. P+L+	1, 15	15.39	37.05	.008**
P+L− vs. P−L−	1, 15	4.66	60.93	.285
P−L+ vs. P−L−	1, 15	0.67	37.53	*ns*
P−L+ vs. P+L−	1, 15	2.38	58.83	*ns*
Midline (4 Conditions × 4 Electrodes)				
Condition	3, 45	9.81	10.58	< .001***
Pairwise comparisons				*p*(corr.)
P+L− vs. P+L+	1, 15	37.72	13.76	< .001***
P−L+ vs. P+L+	1, 15	10.82	28.82	.029*
P−L− vs. P+L+	1, 15	10.82	19.31	.030*
P+L− vs. P−L−	1, 15	2.92	23.76	*ns*
P−L+ vs. P−L−	1, 15	0.75	13.79	*ns*
P−L+ vs. P+L−	1, 15	1.50	17.42	*ns*

Amplitudes of the ERPs were averaged over this time window for every participant separately and entered into repeated measures ANOVA with factors condition (four levels) and quadrant (four levels). A separate ANOVA was performed for the midline electrodes with factors condition (four levels) and electrode (four levels). Huynh–Feldt correction for violation of sphericity assumption was applied (Luck, 2005; Picton et al., 2000; Huynh & Feldt, 1976), but the original degrees of freedom are reported. The significance levels for the pairwise comparisons were corrected for the number of tests performed by means of Bonferroni correction. The corrected *p* levels are reported; effects with *p* values > .5 are reported as not significant (*ns*). P+L+ = correct condition; P+L− = language mismatch condition; P−L+ = picture mismatch condition; P−L− = double mismatch condition.

*p < .05.

**p < .01.

***p < .001.

conditions and we report adjusted *p* values accordingly. In the left anterior quadrant, the language mismatch differed significantly from the correct condition [*F*(1, 15) = 30.74, *MSE* = 15.10, *p* < .001]. The double mismatch condition and the picture mismatch condition did not differ significantly from the correct condition [*F*(1, 15) = 4.31, *MSE* = 40.54, *p* = .33 and *F*(1, 15) = 2.43, *MSE* = 30.47, *p* > .5, respectively]. In the right anterior quadrant, a qualitatively similar pattern of results was observed with language mismatch being significantly different from the correct condition [*F*(1, 15) = 11.54, *MSE* = 32.54, *p* = .024], whereas the picture mismatch

and double mismatch conditions were not different from the correct conditions [*F*(1, 15) = 5.30, *MSE* = 29.86, *p* = .216; P−L− vs. P+L+: *F*(1, 15) = 8.31, *MSE* = 27.53, *p* > .5]. Other comparisons did not reveal significant differences between conditions. Over the midline electrodes, there was only a trend for a main effect of condition [*F*(3, 45) = 2.45, *MSE* = 12.41, *p* = .076].

Discussion

Highly similar N400 effects were found for all experimental conditions compared to the correct condition. Comparing our ERP findings to earlier studies investigating semantic processing of pictures, a few differences are readily apparent. In contrast to some earlier findings (Federmeier & Kutas, 2001, 2002; West & Holcomb, 2002; McPherson & Holcomb, 1999), we failed to observe a separate N300 effect, which has been claimed to be specific to the processing of pictorial stimuli. However, some other studies investigating pictures in a sentence context have also failed to observe a separate N300 effect (Ganis et al., 1996; Nigam et al., 1992). Therefore, we argue for the N300 as not being specific to the processing of pictures, at least not when presented within a sentence context. The absence of a picture-specific effect in the ERP waveforms and the similar time course of the N400 suggest that at this level of processing, no differentiation is made between verbal and visual semantic information.

The double mismatch condition, in which both the word and the picture did fit the previous context less well, evoked an N400 similar in latency and amplitude than the other mismatch conditions, in which either word or picture were in discordance with the previous context. Conflicting information coming from the visual or verbal domain does not add up linearly to increase the effect size in the double mismatch. Furthermore, the fact that the double mismatch N400 starts at the same latency as the other two mismatch conditions speaks in favor of the ''immediacy assumption'' which predicts that information is used by the language comprehension system as soon as it is available. In other words, it suggests that picture and word are not first integrated at a lower level of processing before being integrated into the sentence context. If this were the case, a delay in the N400 response to the language mismatch condition and the picture mismatch condition would have been expected.

Interestingly, an earlier, marginally significant difference between mismatch conditions and the correct condition could be observed in the time window of the P2 component. This is most parsimoniously explained as a lead-in effect of the subsequent N400, which was much more negative for the three mismatch conditions than for the correct condition.

In line with earlier ERP studies, the scalp distribution of the N400 effect was more frontal than the centro-posterior distribution that is normally observed in studies of spoken or written language. The frontal distribution

was, however, not specific to the picture mismatch condition. Therefore, although the presence of visual information might shift the N400 distribution to a more frontal maximum, the fact that this holds even when the anomaly is language-internal argues against a picture-specific integration process that is different from semantic integration of written or spoken words. Together with results in an earlier study (Özyürek et al., 2007), it seems that the mere presence of a visual stimulus (other than a written word) makes the scalp distribution "shift" to a more frontal maximum compared to when only linguistic information is presented.

The stronger negativities to the mismatch conditions in the late time window resemble the findings in the N400 time window, although only the language mismatch condition differed significantly from the correct condition. However, no differences were observed between the experimental conditions. Therefore, these late effects are best explained as a carryover effect of the strong N400 effects.

EXPERIMENT 2: fMRI

Participants

Nineteen healthy right-handed (Oldfield, 1971) participants with Dutch as their mother tongue took part in the fMRI study. None had any known neurological history or hearing complaints, and all had normal or corrected-to-normal vision. Three datasets in the fMRI study had to be discarded, two because of inattentive participants (see below) and one because of excessive head motion. Data from the 16 remaining participants (mean age = 22.3 years, range 20–28, 8 women) were entered into the analysis. Participants were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the Declaration of Helsinki.

Procedure

Stimuli were the same as in the EEG experiment. Pictures were projected from outside of the scanner room onto a screen at the end of the patient table. The screen was visible through a mirror mounted to the head coil, at a viewing distance of 80 cm (pictures subtended maximally 5.7° × 5.7° visual angle). Speech was presented through nonmagnetic headphones (Commander XG, www.mrivideo.com), which dampened scanner noise. Intertrial interval was 6, 7, or 8 sec. During the scanning session eye movements were recorded using an infrared IviewX eyetracker (www.smi.de) to formally control participant's vigilance during scanning.

The scanner was switched on during the practice trials and participants had to indicate whether the volume should go up or down. No participant asked for the volume to be increased to the maximally possible level.

Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterwards. All participants indicated they were able to hear and understand the sentences well. At the end of the scanning session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

Recording and Analysis

MR imaging was performed on a 3-T Siemens Magnetom Trio scanner (Siemens, Erlangen, Germany). Approximately 800 echo-planar whole-brain images were acquired per participant (TR = 2230 msec; TE = 30 msec; flip angle = 80°; 32 slices; slice thickness = 4 mm; FOV = 224 mm, voxel resolution = 3.5 × 3.5 × 4 mm). Additionally, a T1-weighted anatomical scan (3-D MP-RAGE, 192 slices, TR = 2300 msec; TE = 3.93 msec; FOV = 256 mm; slice thickness = 1 mm) was made. Data were analyzed using Brainvoyager QX (Brain Innovation, www.brainvoyager.com). The first five volumes of a session were discarded to avoid T1 saturation effects. Preprocessing involved rigid-body transformations of all volumes to the first volume, slice scan time correction, linear trend removal, and high-pass temporal filtering (cutoff 3 cycles over the time course) and spatial smoothing with a Gaussian filter kernel of 8 mm full width at half maximum. Data were transformed into stereotaxic space (Talairach & Tournoux, 1988). A whole-brain analysis was performed in the context of the General Linear Model, with the conditions as factors of interest and the six parameters from the motion correction as nuisance factors. Experimental factors were modeled for the duration of each sentence and convolved with a canonical 2 gamma hemodynamic response function. Parameters were estimated for every voxel's time course. Effect sizes were estimated by constructing contrast (t) maps consisting of differences between the parameter estimates in every voxel and participant separately for contrasts of interest. Subsequently, contrast maps were taken to a second-level analysis, testing for differences from zero in a one-sample t test in a random effects analysis. Every contrast was tested two-sided. The multiple comparisons problem was addressed by thresholding the activation maps at $t(15) = 3.9$, $p < .001$ at the voxel level and taking the cluster sizes into account, leading to a correction at an alpha level of $p < .05$ (Forman et al., 1995). The eyetracking data were used to control for the vigilance (i.e., wakefulness) of the participant. Two datasets had to be discarded because participants had their eyes closed in more than 10% of the trials.

Results

To see effects specific for the language condition, the language mismatch condition (P+L−) was contrasted against the correct condition (P+L+). The correct con-

dition served as a high-level baseline in this way. An extensive region in the left inferior frontal cortex, stretching into the premotor cortex, and an area in the left superior temporal sulcus were found activated (Table 3; Figure 3A). One area in the right middle frontal sulcus was activated in the reversed contrast (i.e., correct condition vs. language mismatch). Second, effects to the picture condition were assessed, again by comparing it to the correct condition (P−L+ vs. P+L+). Part of the left inferior frontal sulcus showed significant activation to this contrast (Table 3; Figure 3B). Finally, the double mismatch condition compared to correct condition tested the effect of both picture and word being in discordance with the sentence context. This comparison (P−L− vs. P+L+) led to increased activity in an extensive part of the inferior frontal cortex stretching into the premotor cortex, an area in the left superior temporal sulcus, an area in the left temporo-parietal junction, and a small area of activation in the right cerebellum (Table 3; Figure 3C).

Figure 4A displays the results of a conjunction analysis (conjunction as in a logical AND; see Nichols, Brett, Andersson, Wager, & Poline, 2005) testing for overlap between the comparisons described above (P−L+ vs. P+L+ ∩ P+L− vs. P+L+ ∩ P−L− vs. P+L+). One region in the left inferior frontal cortex (max TAL [−40 11 31]) was found activated in this contrast. Compared to a cytoarchitectonic probability map, 39% of this region overlapped with BA 44, in contrast to only 3% of the voxels that were classified as being part of BA 45 (Eickhoff et al., 2005).

Finally, we tested for differential effects to the local and global match or mismatch effects. This was done by

comparing the language mismatch and picture mismatch to the double mismatch condition (P+L− vs. P−L− ∩ P−L+ vs. P−L−). In this way, all conditions involve a sentence-level mismatch, but only the language and picture mismatch conditions had an additional local mismatch. No areas were found activated in this contrast.

Discussion

We observed increased activation levels in all three mismatch conditions compared to the correct condition in the left inferior frontal cortex. This study adds to a large number of studies showing that the left inferior frontal cortex is an important node in the speech comprehension network (for reviews, see Vigneau et al., 2006; Bookheimer, 2002). We interpret our findings as reflecting unification processing in the left inferior frontal cortex. This entails integration of information into a built-up representation of the previous sentence context as well as a selection of appropriate candidates for integration (Hagoort, 2005a, 2005b). We show here that integration of extralinguistic information, such as a visual picture, recruits this area.

The peak of overlapping activation in the inferior frontal cortex was in BA 44, which is at odds with proposals of a gradient of linguistic subprocesses (such as semantics, syntax, and phonology) within the left inferior frontal cortex (Vigneau et al., 2006; Bookheimer, 2002). Semantic processing is centered more ventrally, around BA 45 and 47 in these proposals (Vigneau et al., 2006; Bookheimer, 2002). However, as can be seen in Figure 3, two of the mismatch conditions did activate a more ventral part of the left inferior frontal cortex (language

Table 3. Results from the fMRI Experiment

Contrast	Center Coordinates			Region	t (Max)	Number of Voxels (1 × 1 × 1 mm)
	x	y	z			
Language mismatch versus correct condition	−45	15	24	Left inferior frontal sulcus/premotor cortex	8.25	8655
	−52	−34	−1	Left superior temporal sulcus	5.42	2507
Picture mismatch versus correct condition	−38	8	32	Left inferior frontal sulcus	6.23	2404
Double mismatch versus correct condition	−43	17	23	Left inferior frontal sulcus/premotor cortex	11.44	17304
	−50	−29	−6	Left superior temporal sulcus	11.20	9370
	−38	−60	14	Left temporo-parietal junction	5.84	3103
	17	−38	−27	Right cerebellum	5.46	906
Correct condition versus language mismatch	28	30	44	Right middle frontal sulcus	7.15	1965

Regions are presented that were significantly activated in the whole-brain random effects group analysis [t(15) > 3.9, p < .05, corrected] by contrasting each mismatch condition against the correct condition. Displayed are the contrasts, the center coordinates in stereotaxic space (Talairach & Tournoux, 1988), a description of the region, the t value of the maximally activated voxel, and the number of significant voxels.
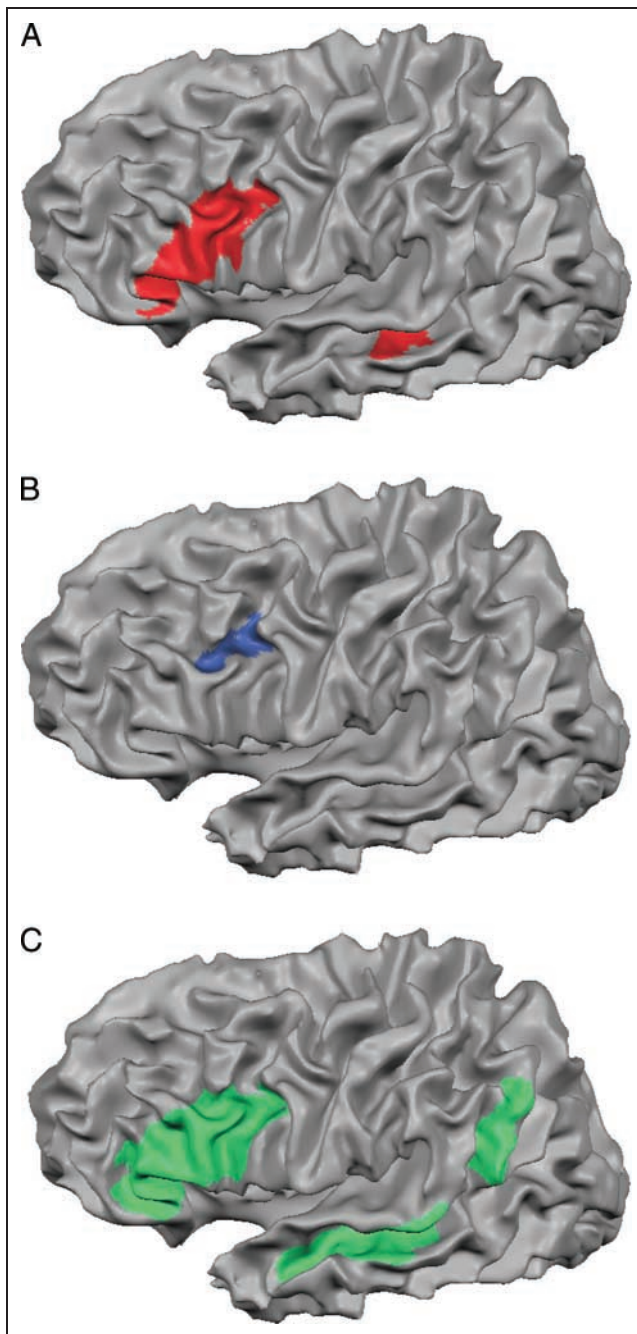
be taken as evidence in favor of or against a gradient of linguistic processes in the left inferior frontal cortex.

A potential worry is that the activation of the left inferior frontal cortex could be a by-product of using the mismatch paradigm, bearing little relevance to general language processing. As is clear from Figure 4B, however, the correct condition also resulted in an activation increase in this region. The activation levels of the correct condition were found to be significantly different from zero (see Figure 4). We show that the inferior frontal cortex activation is involved also in the processing of semantically correct sentences and that its activation in this study cannot be attributed to the use of the mismatch paradigm (see also Hasson, Nusbaum, & Small, 2007; Willems et al., 2007; Hagoort, 2005b; Rodd et al., 2005).

In relation to the object priming literature cited in the Introduction, we want to point out that our results cannot be explained in terms of increased conceptual priming in the correct condition. That is, word and picture were conceptually the same in both the correct condition and the double mismatch condition. However, increased inferior frontal cortex activation was,
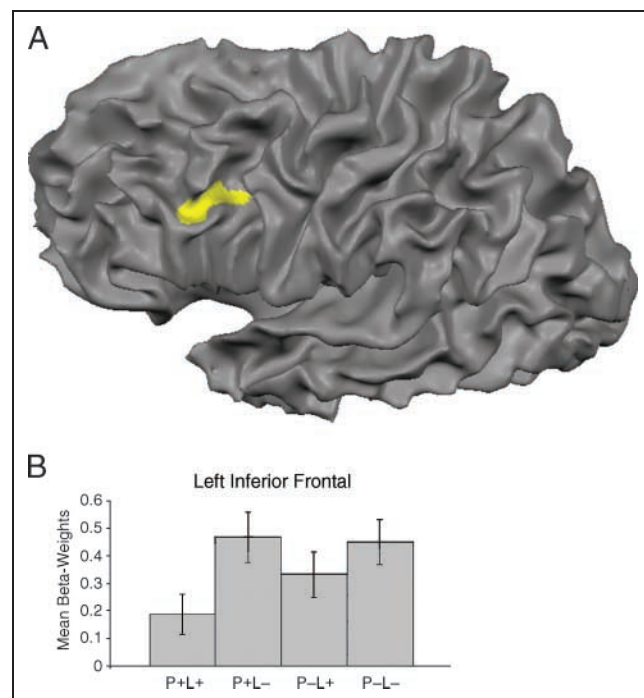
Figure 3. Results from the fMRI whole-brain random effects group analysis [$t(15) > 3.9$, $p < .05$, corrected]. Areas significantly activated in the (A) language mismatch versus correct condition contrast (red), (B) picture mismatch versus correct condition contrast (blue), (C) double mismatch versus correct condition contrast (green). Results are overlain on a cortical sheet segmented along the gray–white matter border in stereotaxic (Talairach) space.

mismatch condition and double mismatch condition). More in general, it should be pointed out that what is most striking from meta-analyses (such as Vigneau et al., 2006; Bookheimer, 2002) is the large spread around the mean center coordinates. Because of this variance across studies, one single study such as the present one cannot

Figure 4. (A) Part of the left inferior frontal cortex commonly activated by all mismatch conditions. This area was the result of a conjunction analysis (Nichols et al., 2005) of each mismatch condition to the correct condition (P−L+ vs. P+L+ ∩ P+L− vs. P+L+ ∩ P−L− vs. P+L+). (B) Parameter estimates (compared to implicit baseline) for all conditions from the area in A. Although this area is more strongly activated in all mismatch conditions than in the correct condition, the parameter estimate to the correct condition was significantly different from zero [$t(15) = 2.43$, $p < .03$], indicating that the area is also activated in the correct condition. Error bars indicate standard error of the mean (SEM).

nevertheless, observed in the double mismatch condition. The reason that no priming effects are observed is probably because picture and word are not presented after another as is usually done in priming paradigms. That is, a picture or a word does not form the context for the other item; rather, the preceding sentence is the crucial context.

Finally, apart from overlapping areas across conditions, we also found an increase in activation in the left superior temporal sulcus specific to the language and double mismatch conditions, but not to the picture mismatch condition.[1] A similar result was obtained in an earlier study investigating the processing of co-speech gestures and words (Willems et al., 2007). This suggests that superior temporal regions might be specifically involved in verbal semantics.[2] Interestingly, no specific effect was observed for the picture mismatch condition. Given the role of the ventral temporal cortex in object representations, this area might have been expected to be more activated in the picture mismatch condition as compared to the other conditions. Such effect was not observed however.

GENERAL DISCUSSION

In this study, we compared the integration of semantic information conveyed through spoken language (words) and visual information (pictures) at the sentence level. Overall, our results provide strong evidence for both processes to tax the same neural processes. That is, neural indicators of semantic integration react the same to both a higher integration load when information is conveyed through a word than when it is conveyed through a picture. A same neural time course is indicated by same onset latencies and effect sizes of the N400 effects. The processing at this level of comprehension does not give temporal precedence to linguistic information over extralinguistic information as indicated by the N400 effects. In terms of neural locus, part of the left inferior frontal cortex was commonly activated by all mismatch conditions. A recent neurobiological account of sentence comprehension has interpreted increased activation in the left inferior frontal cortex as being the neural indicator of increased integration load of a word's meaning into a built-up (sentence) context (Hagoort, 2005b). Here we provide evidence for this region not to be domain-specific because the integration of information presented in a nonlinguistic modality also taxes this region. This is in line with an earlier study in which we found the left inferior frontal cortex to be activated more strongly to both spoken words and co-speech gestures in a sentence context (Willems et al., 2007). However, there is an important difference between our previous and the present study. Co-speech gestures are necessarily bound to a language context; that is, they do not clearly represent their meaning when presented on their own (McNeill, 1992; Krauss et al., 1991). Pictures,

on the contrary, are fully meaningful outside of a language context. In this way, the present study provides more convincing evidence for the claim that the role of the left inferior frontal cortex in language comprehension is not domain-specific. The left inferior frontal cortex plays an important role in integration and selection operations that combine linguistic and extralinguistic visual information into a coherent overall interpretation of an expression.

The current study adds to an understanding of the language comprehension system as taking several types of information into account in the same way when understanding a message (see also Spivey Knowlton & Sedivy, 1995; Tanenhaus & Trueswell, 1995; Trueswell & Tanenhaus, 1994; Taraban & McClelland, 1990). That is, the system does not restrict itself to one source of information (speech), but seems to use a rich variety of sources of meaningful information in a qualitatively similar way when understanding a message. Note that the visual information in our study was rather simple, consisting of pictures of single objects. In contrast, in the eye movement literature cited above, visual context often involves several objects (e.g., Tanenhaus et al., 1995). For reasons of comparability with earlier studies (e.g., Özyürek et al., 2007) as well as for reasons of experimental control, we restricted ourselves to using pictures of single objects as stimuli. However, the few ERP studies that did use a richer visual context seem to suggest that similar findings would be obtained if the visual stream of information had been richer (e.g., Ganis & Kutas, 2003; Sitnikova et al., 2003; West & Holcomb, 2002). This is, however, an issue that is open for empirical investigation.

Importantly, our results support a theory of language processing that goes against the classical two-step model of interpretation (e.g., Lattner & Friederici, 2003; Cutler & Clifton, 1999). Instead, in line with the immediacy assumption, all available information is used directly to co-determine the interpretation of linguistic expressions. Moreover, we show that the role of the inferior frontal cortex in the language comprehension network is not restricted to linguistic information. Rather, also an increased semantic integration load conveyed by a picture activates this area.

Notes

1. Informal visual inspection of the contrast map at a lower statistical threshold confirmed that in the picture mismatch–correct condition comparison, the superior temporal cortex was not activated.

2. This could relate to the debated issue of the semantic system to be organized in multiple semantic codes or in one common code (Caramazza, Hillis, Rapp, & Romani, 1990; McCarthy & Warrington, 1988; Shallice, 1988). Our study was designed to investigate sentence-level integration of semantic information and not targeted at revealing the neural representations of words and pictures. Therefore, we are reluctant to interpret our results in terms of multiple or modality-specific codes.

REFERENCES

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14,* 201–212.

Baumgaertner, A., Weiller, C., & Buchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage, 16,* 736–745.

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

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

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology, 7,* 161–189.

Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.

Culicover, P. W., & Jackendoff, R. (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences, 10,* 413–418.

Cutler, A., & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language*. Oxford: Oxford University Press.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 16032–16037.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25,* 1325–1335.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 202–224.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40,* 730–747.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33,* 636–647.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & C. T. Walker (Eds.), *Sentence processing: Psycholinguistic essays presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13,* 170–177.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16,* 123–144.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for ''common sense'': An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8,* 89–106.

Hagoort, P. (2005a). Broca's complex as the unification space for language. In A. Cutler (Ed.), *Twenty first century psycholinguistics: Four cornerstones* (pp. 157–172). Mahwah, NJ: Erlbaum.

Hagoort, P. (2005b). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9,* 416–423.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In L. Frazier, J. Clifton Charles, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 45–80). Hillsdale, NJ: Erlbaum.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–441.

Hagoort, P., & van Berkum, J. J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society, Series B, 362,* 801–811.

Hasson, U., Nusbaum, H. C., & Small, S. L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. *Cerebral Cortex, 17,* 2899–2913.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24,* 259–276.

Huynh, H., & Feldt, L. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and splitplot designs. *Journal of Educational Statistics, 1,* 69–82.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7,* 302–307.

Krauss, R. M., Morrel Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61,* 743–754.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15,* 272–293.

Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., et al. (2000). Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *Journal of Cognitive Neuroscience, 12,* 321–341.

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

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

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

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing-evidence from event-related brain potentials. *Neuroscience Letters, 339,* 191–194.

Lebreton, K., Desgranges, B., Landeau, B., Baron, J. C., & Eustache, F. (2001). Visual priming within and across symbolic format using a tachistoscopic picture identification task: A PET study. *Journal of Cognitive Neuroscience, 13,* 670–686.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge: MIT Press.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11,* 194–201.

McCarthy, R. A., & Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. *Nature, 334,* 428–430.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McPherson, W., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36,* 53–65.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology, 35,* 99–115.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12,* 120–133.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653–660.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4,* 15–22.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

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. *Journal of Cognitive Neuroscience, 19,* 605–616.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37,* 127–152.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15,* 1261–1269.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron, 20,* 185–195.

Shallice, T. (1988). Specialisation within the semantic system. *Cognitive Neuropsychology, 5,* 133–142.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40,* 160–164.

Spivey Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55,* 227–267.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

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

Taraban, R., & McClelland, J. L. (1990). Parsing and comprehension: A multiple-constraint view. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 231–263). Hillsdale, NJ: Erlbaum.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A., 94,* 14792–14797.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, Jr. (Ed.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

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. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443–467.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia, 33,* 485–508.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage, 30,* 1414–1432.

Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., & Gabrieli, J. D. (1997). Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *Journal of Cognitive Neuroscience, 9,* 714–726.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13,* 363–375.

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