Event-related Potential Evidence of Form and Meaning Coding during Online Speech Recognition

Claudia K. Friedrich[1] and Sonja A. Kotz[2]

Abstract

■ It is still a matter of debate whether initial analysis of speech is independent of contextual influences or whether meaning can modulate word activation directly. Utilizing event-related brain potentials (ERPs), we tested the neural correlates of speech recognition by presenting sentences that ended with incomplete words, such as *To light up the dark she needed her can-*. Immediately following the incomplete words, subjects saw visual words that (i) matched form and meaning, such as *candle*; (ii) matched meaning but not form, such as *lantern*; (iii) matched form but not meaning, such as *candy*; or (iv) mismatched form and meaning, such as *number*. We report ERP evidence for two distinct cohorts of lexical tokens: (a) a left-lateralized effect, the P250, differentiates form-matching words (i, iii) and form-mismatching words (ii, iv); (b) a right-lateralized effect, the P220, differentiates words that match in form and/or meaning (i, ii, iii) from mismatching words (iv). Lastly, fully matching words (i) reduce the amplitude of the N400. These results accommodate bottom–up and top–down accounts of human speech recognition. They suggest that neural representations of form and meaning are activated independently early on and are integrated at a later stage during sentence comprehension. ■

INTRODUCTION

Successful speech recognition depends not only on the identification of individual word meaning, but also on meaning conveyed by word combinations as in a sentence. Word identification serves as an interface between the speech input and the construction of an interpretative representation of the speaker's utterance (Zwitserlood, 1999). Word identification processes involve the activation of a cohort (Marslen-Wilson & Welsh, 1978) of lexical representations that are temporally or partially consistent with the speech signal (Spivey, Grosjean, & Knoblich, 2005; Allopenna, Magnuson, & Tanenhaus, 1998; Zwitserlood, 1989; Marslen-Wilson, 1987, 1989). At present, there is a controversy on whether the ongoing sentence interpretation can modulate cohort activation of subsequently presented words.

Theories of cognition can be divided into two categories: *Modular theories* propose perceptual analysis independent of contextual influences. Here, integration of different types of information is accomplished during late processing stages (Fodor, 1983; Forster, 1979). *Interactive theories*, on the other hand, promote that multiple sources are combined during perceptual analysis (Grossberg & Myers, 2000; McClelland, 1987). In the field of speech recognition, this distinction divides theoretical accounts of information flow (see Bowers & Davis, 2004; Cutler & Clifton, 2000; Zwitserlood, 1999, for recent reviews). In modular models of speech processing, activation of word information stored in memory spreads bottom–up with no information flow in the opposite direction. Speech sounds activate words and words activate meaning representations (Norris, McQueen, & Cutler, 2000; Marslen-Wilson, 1987, 1989). In interactive or distributed models of speech processing, information flows in both directions. Words receive activation from both bottom–up and top–down mechanisms (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986).

Evidence for an Initially Form-based Cohort

Empirical evidence in favor of modular models comes from behavioral cross-modal priming studies. Subjects listened to spoken sentences and were asked to decide whether an occasionally presented visual letter string is a word or not. The majority of these behavioral cross-modal priming studies used ambiguous words. By and large, the results show that immediately following ambiguous words, all meanings are activated even if prior context biases one meaning. For example, directly after the offset of the word *bank* spoken in a financial context, facilitation is found for the words *money* and *river* (Onifer & Swinney, 1981; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979; but see Tabossi & Zardon, 1993).

Perhaps the strongest behavioral support for context-independent cohort activation comes from sentence

[1]University of Hamburg, [2]Max Planck Institute for Human Cognitive and Brain Sciences, Germany

fragment priming. Zwitserlood (1989) presented ambiguous sentence final word fragments, such as *dol-*, that can be completed by more than one word, such as *dollar* or *dolphin*. Crucially, targets related to appropriate word completions that did not fit the sentence meaning were facilitated similarly to targets related to appropriate word completions that did fit the sentence meaning. That is, *dol-* facilitated behavioral responses to *dolphin* even if the sentence context biased a ''financial'' concept. Zwitserlood's results demonstrate that an initial cohort is activated based on word form regardless of context. Comparable results are found for priming of ambiguous word onsets. For example, a string with two potential interpretations facilitated associations of both words even in a biasing sentence such as *She mended the d/tuck* (Coninne, Blasko, & Wang, 1994). Taken together, facilitated lexical decision times for semantic associates of ambiguous words and fragments support initial cohort activation independent of context. However, these studies tested activation at a semantic level of representation. Thus, the results may not be generalizable for initial word form activation. Furthermore, reaction times only reflect the outcome of a complex decision process and, therefore, may not effectively capture rapid activation patterns during online speech recognition.

Evidence for Immediate Incremental Processes

Online signatures of semantic processing in the event-related brain potential (ERP) challenge the assumption of delayed contextual influence. The N400, a negative peaking component (400 msec following word onset) is correlated with meaning integration into a preceding context (Kutas & Hillyard, 1984; see Kutas & Federmeier, 2000, for reviews). The onset of the N400 reveals that mismatching sentence final words that share an initial syllable with an appropriate sentence completion reduce the semantic integration effort before the acoustic signal allows unique word identification (Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Sentences such as *It was a pleasant surprise to find that the car repair bill was only seventeen*... ended with a congruent word like *dollars*, with an onset congruent word like *dolphins*, or with an onset incongruous word like *scholars*. Both congruent and onset congruent words initially showed reduced N400 amplitudes compared with incongruous words. Similarly, negative peaking components preceding the N400, alternatively labeled as phonological mismatch negativity (Connolly & Phillips, 1994; Connolly, Phillips, Steward, & Brake, 1992) or N200 (van den Brink, Brown, & Hagoort, 2001; Hagoort & Brown, 2000), may reflect initially reduced integration effort for words with similar word onsets as ''cloze'' words in an ongoing sentence. Together these findings indicate that partial information from perceptual analysis is integrated with contextual information as soon as it becomes available.

Recent eye movement evidence shows that word identification is immediately influenced by sentence context (Dahan & Tanenhaus, 2004). Participants identified referents of subjects of spoken Dutch sentences among visual distractors. Crucially, verb-based semantic constraints differentially modulated the number of potentially fitting subjects. For example, the verb in the sentence *Nog nooit klom een bok zo hoog* (''Never before climbed a goat so high'') places strong semantic constraints on its subject, whereas the verb preceding the subject in the sentence *Nog nooit is een bok zo hoog geklommen* (''Never before has a goat climbed so high'') does not place such constraints. Only in sentences with low semantic constraints were eye movements attracted toward phonological competitors of the subject (e.g., *bot* [bone]). Thus, in contrast to lexical decision latencies (Zwitserlood, 1989), eye movements suggest that context is immediately used to restrict the cohort activated by the incoming signal. Opposite effects in both studies may represent different activation patterns or processing mechanisms used in lexical decisions or the preparation of eye movements. In order to disentangle these effects, an online analysis of parallel activation during cognitive processing is needed.

The Present Study

In the current study, we tested sentence fragment priming while recording ERPs to track subprocesses during lexical access. Previously we have shown that the presentation of spoken word fragments followed by complete visual words elicits a specific left-hemispheric ERP effect. Contrasting the ERPs for words that match preceding fragments (e.g., *mus- museum*) with unrelated words (e.g., *fac- museum*) results in a positivity between 200 and 400 msec with a peak at 350 msec, called the P350 effect (Friedrich, 2005; Friedrich, Kotz, Friederici, & Alter, 2004; Friedrich, Kotz, Friederici, & Gunter, 2004). In line with previous psycholinguistic research (Gaskell & Marslen-Wilson, 2002; Cutler & van Donselaar, 2001; Soto-Faraco, Sebastian-Galles, & Cutler, 2001), we related this neurophysiological effect to the activation of a cohort of lexical representations that match preceding fragments. Time course and source localization within left temporal brain regions suggest a link between the P350 effect and a magnetoencephalographic response, the M350, which has also been linked to lexical activation (Pylkkänen, Llinás, & Murphy, 2006; Pylkkänen & Marantz, 2003; Pylkkänen, Stringfellow, & Marantz, 2002).

In the present study, we aimed to elucidate whether cohort activation is exclusively related to word form (bottom–up) or can be modulated by meaning (top–down). Comparable to Zwitserlood (1989), we presented sentences that ended with ambiguous fragments, such as *To light up the dark she needed her can-*. Immediately following the fragments, we presented visual probes in order to test whether different types of lexical representations are activated by the preceding sentence fragments:

(i) In a *complete match* condition, we tested representations that matched both the form of the ambiguous fragment and the meaning of the sentence, such as candle.
(ii) In a *meaning match* condition, we tested activation of representations that did not match the form of the ambiguous fragment, but the meaning of the sentence, such as lantern.
(iii) In a *form match* condition, we tested representations that matched only the form of the ambiguous fragment, but did not match the meaning of the sentence, such as candy.
(iv) In an *unrelated* condition, we tested responses to representations that matched neither the form of the fragment nor the meaning of the sentence, such as number (see Figure 1A for German examples presented in the experiment).

We tested the following accounts of lexical identification:

(A) *Form-based activation.* Following a purely bottom–up approach to early phases of lexical identification, sentence final fragments should only activate form-matching words. Thus, complete match (i) and form match (iii) should elicit comparable ERP waveforms that differ from meaning match (ii) and mismatch (iv).
(B) *Meaning-based activation.* If a specific neural network is mainly sensitive to sentence context, ERPs for complete match (i) and meaning match (ii) on the one hand, and ERPs for form match (iii) and mismatch (iv) on the other, should be contrastable.
(C) *Form-and-meaning-based activation.* If both form and context influence early stages of lexical identification, form-based and context-based representations

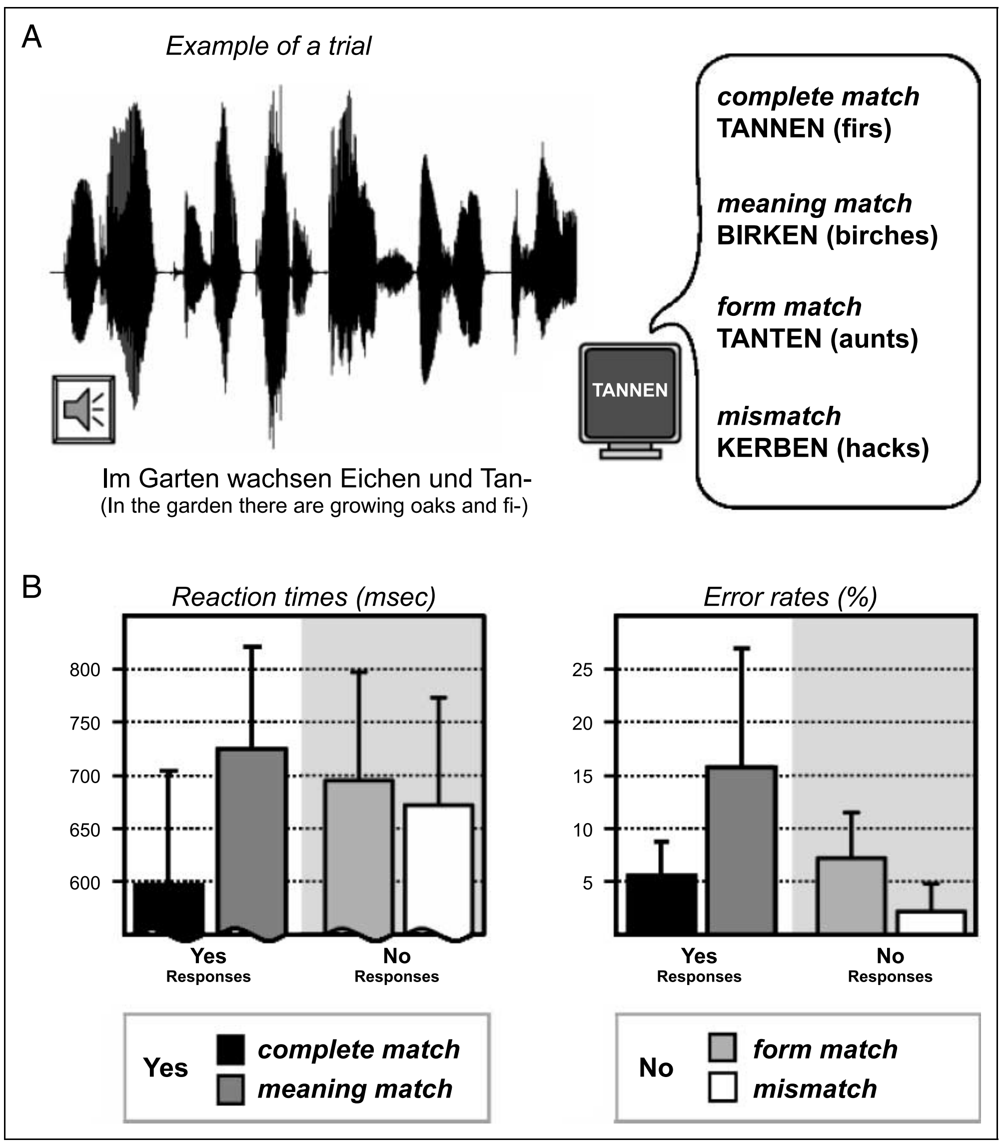

Figure 1. Design of the present study and behavioral responses. (A) Subjects listened to spoken sentences with incomplete final words. Visual words that immediately followed the fragmented end matched sentence meaning and form of the ambiguous fragment (i), matched meaning but not form (ii), matched form but not meaning (iii), mismatched both (iv). (B) Reaction times and error rates for the decisions whether or not the visual word matched the sentence meaning, regardless of the incomplete word: (i) black bars, (ii) dark gray bars, (iii) light gray bars, (iv) white bars.

should be activated. Consequently, all types of match (i)–(iii) should elicit similar ERPs that differ from unrelated words (iv).

(D) *Integration of form and meaning.* An early interaction of form and meaning information would favor words that match both the word fragment and the sentence context. If a complete match (i) differs from the other conditions (ii) to (iv), this should support a strong interactive account of lexical activation.

METHODS

Participants

Data were collected from 24 subjects (12 women, 12 men) aged 19 to 27 years (mean, 22.4 years). All participants were native speakers of German with no discernible uncorrected deficits in hearing or vision. Participants were paid for their participation. Only right-handers, as ascertained by the Edinburgh Handedness Questionnaire (Oldfield, 1971), were included.

Stimuli

Initially, four versions of 60 quadruplet sentences were constructed. Two versions of each sentence had congruous final words (i, ii), for example, (i) *Im Garten wachsen Eichen und Tannen* ("In the garden there are growing oaks and firs") and (ii) *Im Garten wachsen Eichen und Birken* ("In the garden there are growing oaks and birches"). In the third version (iii), the sentence final word shared the first syllable with the final word in version (i), but did not match the sentence context, for example, *Im Garten wachsen Eichen und Tanten* ("In the garden there are growing oaks and aunts"). The fourth version of each sentence (iv) had a final word that neither matched the sentence context nor the onset of a sentence congruent word (iv), for example, *Im Garten wachsen Eichen und Kerben* ("In the garden there are growing oaks and hacks"). All sentence final words were disyllabic nouns with first-syllable stress.

A behavioral experiment was conducted to select items for the ERP priming experiment. A questionnaire containing all 240 sentences in random order was presented to 20 participants who did not participate in the following ERP experiment. Participants decided on a 5-point scale whether the final word fits the respective sentence or not. The median number of subjects who decided that the final word fits (congruous median, including rankings of 4 and 5) and the median number of subjects who decided that the final word did not fit (incongruous median, including rankings of 1 and 2) were determined for each sentence.

A total of 40 quadruplet sentences were selected according to three criteria. First, both types of congruous words that represent complete match and meaning match had to be judged as equally congruous sentence endings. Accordingly, congruous medians for complete match (9.3 subjects) and for meaning match (9.1 subjects) did not differ for the final set of sentences, $t(19) = 0.87$, *ns*. Second, both types of semantically incongruous words (iii) and (iv) had to be judged as equally poor sentence endings. Thus, incongruous medians for form match (8.6 subjects) and mismatch (8.9 subjects) were comparably high, $t(19) = 2.46$, *ns*. Congruous and incongruous medians for complete match and meaning match differed from congruous and incongruous medians for form match and mismatch (all $p < .001$). Finally, word length and frequency of words presented in the different conditions were balanced using an online dictionary of German (wortschatz.uni-leipzig.de).

Only versions of sentences with completely matching final words (i) were spoken by a female native speaker of German. Stimuli were recorded with a DAT recorder at a sampling rate of 44.1 kHz and then transferred to a computer, volume equalized, and edited into individual tokens using the Cooledit 2000 (www.adobe.com/special/products/audition/syntrillium.html) waveform manipulation software package. The last syllable of the sentence final word was removed from the respective auditory file. This resulted in 40 prime sentences with incomplete final words. Each prime was used in four different conditions (see Figure 1A): complete match (i); meaning match (ii); form match (iii); and mismatch (iv). Each of four experimental blocks contained one presentation of each sentence with one of the four different targets. The order of targets was pseudorandomized with the constraint that no more than three "yes" responses or three "no" responses followed directly. The order of blocks was counterbalanced across participants.

Procedure

Participants sat in a sound-attenuated booth. Prime stimuli were presented via loudspeakers. Each trial began with the auditory presentation of an incomplete sentence immediately followed by the visual presentation of a target word. Targets were presented for 300 msec (see Figure 1A for illustration). Participants had to indicate whether the target word matched the sentence meaning while ignoring the sentence final fragment of the auditory primes. Speed and accuracy were equally stressed. Half the participants made yes responses with the thumb of their left hand and no responses with the thumb of their right hand. For the remaining participants, response hands were reversed. A break was introduced after half of the trials.

Ag/AgCl electrodes were held in place on the scalp with an elastic cap. Scalp locations included 60 standard international 10-10 system locations (Chatrian, Lettich, & Nelson, 1988). To control for eye movements, two electrodes were placed below and above the left eye and two electrodes were placed beside both eyes. All electrodes were referenced to the nose. All electrode impedances were less than 5 kΩ. The electroencephalogram (EEG) was recorded with a sampling rate of 250 Hz.

Data Analyses

Error rates were calculated for all words. Reaction times and ERPs were calculated for correct responses only. A two-way analysis of variance (ANOVA) with the repeated measures factors Form (two levels): complete match (i) and form match (iii) versus meaning match (ii) and mismatch (iv); and Meaning (two levels): complete match (i) and meaning match (ii) versus form match (iii) and mismatch (iv), was applied to analyze reaction times and error rates separately. Because participants judged whether or not the visually presented target word matched the sentence meaning, response categories are similar for all levels of the factor meaning (see also Figure 1B). Because different strategies may underlie the rejection or acceptance of a trial, follow-up comparisons were conducted separately for the yes-response condition and for the no-response condition.

For EEG analysis, trials were averaged off-line with an epoch length of 800 msec, including a prestimulus baseline from −200 to 0 msec (with respect to the onset of the visual word). Six lateral electrode groups, each composed of eight electrode positions, were defined in accordance with cortical regions: left frontal (FP1, AF7, AF3, F9, F7, F5, F3, FT9); left temporocentral (FT7, FC5, FC3, T7, C5, C3, CP5, CP3); left parietooccipital (TP9, TP7, P9, P7, P5, P3, PO7, PO3); right frontal (FP2, AF4, AF8, F4, F6, F8, F10, FT10); right temporocentral (FC4, FC6, FT8, C4, C6, T8, CP4, CP6); right parietooccipital (TP8, TP10, P4, P6, P8, P10, PO4, PO8). Four-way ANOVAs with the repeated measures factors Form (two levels): complete match (i) and form match (iii) versus meaning match (ii) and mismatch (iv); Meaning (two levels): complete match (i) and meaning match (ii) versus form match (iii) and mismatch (iv); Hemisphere (two levels): left electrode leads versus right electrode leads; and Region (three levels): frontal electrode leads versus temporocentral electrode leads versus parietooccipital electrode leads, were applied to compare mean ERP amplitudes of the experimental conditions. Time windows that showed condition effects were identified using 50-msec time step ANOVAs (see Table 1). For illustration purposes only, ERPs were low-pass filtered at 10 Hz.

We tested the four hypotheses specified in the Introduction separately. Based on hypothesis (A) (form-based activation), main effects of factor Form were expected. Based on hypothesis (B) (meaning-based activation), main effects of factor Meaning were expected. Confirming hypotheses (C) and (D) required significant interactions of Form and Meaning. By weighting one condition (parameter −1) against the remaining conditions (each with parameter 1), two different models were explored by means of contrast analyses. Hypothesis (C) (form-and-meaning-based activation) was tested by means of a model that contrasted unrelated words (iv) against the other conditions (i) to (iii). Hypothesis (D) (integration of form and meaning) was tested by means of a model that contrasted completely matching words (i) against the other conditions (ii) to (iv).

RESULTS

Behavioral Responses

ANOVAs revealed significant interactions of factors Meaning and Form for reaction times, $F(1,23) = 161.46$, $p <$

Table 1. Results for 50-msec Time Step ANOVAs (See Methods Section) Beginning with Target Onset up to 600 msec Thereafter

	0–50 msec	*50–100 msec*	*100–150 msec*	*150–200 msec*	*200–250 msec*	*250–300 msec*	*300–350 msec*	*350–400 msec*	*400–450 msec*	*450–500 msec*	*500–550 msec*	*550–600 msec*
Form × Hemisphere						X	X					
Form × Region									X	X		
Form × Hemisphere × Region									X			
Meaning × Hemisphere								X				
Meaning × Region										X	X	
Meaning × Hemisphere × Region												
Form × Meaning								X	X	X	X	X
Form × Meaning × Hemisphere					X		X					
Form × Meaning × Region									X	X	X	X

Significant interactions ($p = .01$) including the factors Form and/or Meaning are marked with an X.

.001, and for error rates, $F(1,23) = 26.87, p < .001$. "Yes" decisions to meaning match were slower, $t(23) = 205.76$, $p < .001$, and more erroneous, $t(23) = 16.05, p < .001$, than "yes" decisions to complete match. "No" decisions to form match were slower, $t(23) = 5.82, p < .03$, and more erroneous, $t(23) = 24.97, p < .001$, than "no" decisions to mismatch (see Figure 1B). These results suggest that subjects had difficulties in ignoring the incomplete word. Decisions were harder in conditions in which form and meaning information conflicted than in conditions in which both types of information did not interact.

Event-related Potentials

Figure 2 illustrates ERPs observed in the present experiment. At around 200 msec, ERPs differ between the visual word conditions. Between 200 and 300 msec, ERPs show a prominent correlate for form overlap over left-hemisphere leads (see Figure 2). Complete match (i) and form match (iii) diverge from meaning match (ii) and unrelated words (iv). This left-hemispheric ERP deflection related to form is confirmed by interactions including factors Form and Hemisphere in two consecutive 50-msec time windows (200–250 and 250–300 msec; see Table 1). A respective time window (200–300 msec) was chosen to analyze the left-hemispheric ERP effect related to word form information.

Within the same time frame, ERPs elicited over right-hemisphere leads were sensitive to both form match and meaning match. Between 200 and 250 msec, a similar pattern for complete match, meaning match, and form match was observed (see Figure 2) and confirmed by an interaction of the factors Form, Meaning, and Hemisphere (see Table 1). Consequently, this time window was chosen to further analyze the ERP deflection with right-hemispheric topography.

Starting at 300 msec a negative-going component, the N400, developed. The N400 is reduced for completely matching words (i) compared to the other three conditions (ii)–(iv). Fifty-millisecond time step ANOVAs confirmed that factors Form and Meaning interact in all time windows 300 msec after target word onset. Accordingly, the N400 effect was further analyzed in a time window ranging from 300 to 600 msec.

Form-based Activation

A four-way ANOVA of the time window between 200 and 300 msec revealed significant interactions of factors Form and Hemisphere, $F(1,23) = 11.63, p = .01$, and Form, Meaning and Hemisphere, $F(1,23) = 6.27$, $p = .02$. Two-way ANOVAs including Form and Meaning revealed a significant effect of Form for the left hemisphere, $F(1,23) = 15.48, p = .001$. The word form effect reflects a positive deflection to words (i) and (iii) that match the fragment as compared with words (ii) and (iv) that do not match the fragment. Neither a main effect of meaning, $F(1,23) < 1$, nor an interaction of form and meaning, $F(1,23) < 1$, was observed for the left hemisphere, indicating that the left-hemispheric ERP deflection is mainly sensitive to word form activation. Taken together, the P250 effect confirms assumption (A) (form-based activation).

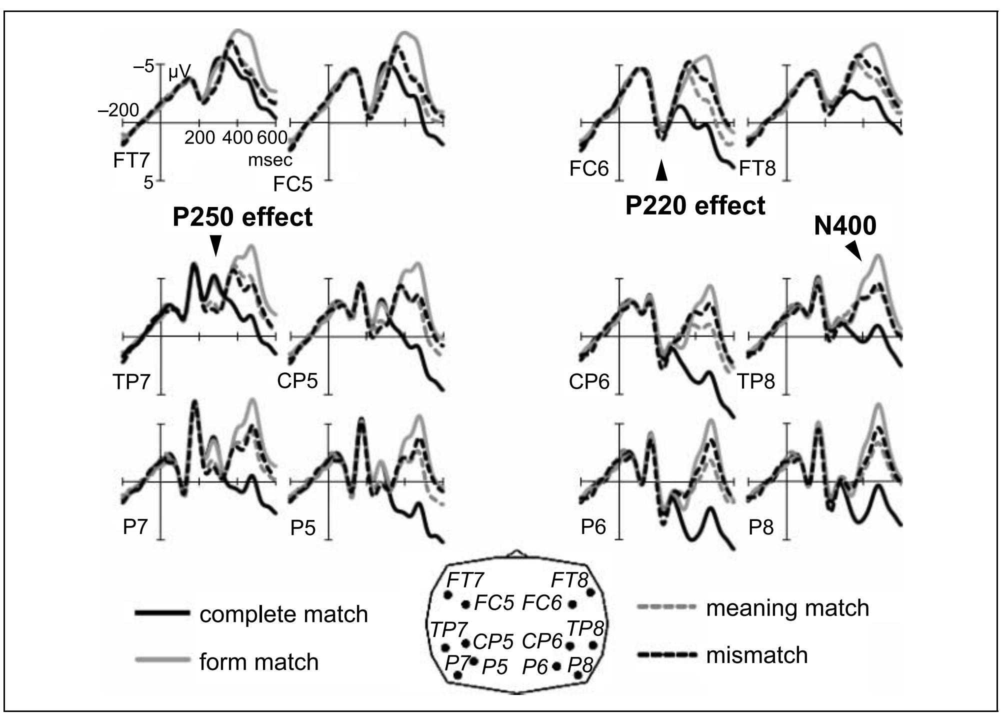

Figure 2. ERPs related to form and meaning processing in speech comprehension: complete match (black solid lines), form match (gray solid lines), meaning match (gray dashed lines), and mismatch (black dashed lines). Starting at 200 msec, ERPs differentiated matching and mismatching words. Left-hemisphere electrode leads: correlates of word form activation (P250 effect). Right correlates of form-and-meaning activation (P220 effect). The respective P250 and P220 effects are further illustrated in Figure 3. Integration of form and meaning information is reflected in the N400 component.

Subtracting form-matching words (i) and (iii) from mismatching words (iv) revealed characteristic positive-going difference waves that started at 200 msec and peaked at 250 msec over the left hemisphere, with a maximum at temporoparietal leads. This effect is henceforth referred to as the P250 difference wave (see Figure 3).

Form-and-Meaning-based Activation

A significant interaction of factors Form, Meaning, and Hemisphere, $F(2,46) = 7.36$, $p = .01$, between 200 and 250 msec was found. Contrast analysis for the right hemisphere suggests that unrelated words (iv) differed from the other conditions (i)–(iii), Wilks' lambda adjusted $F(3,23) = 2.86$, $p = .06$. Post hoc comparisons for right-hemispheric electrode leads confirmed that mean amplitudes for unrelated words differed from the remaining conditions: (i) completely matching words, $t(23) = 4.39$, $p = .05$; (ii) meaning-matching words, $t(23) = 3.45$, $p = .08$; and (iii) form-matching words, $t(23) = 6.44$, $p = .02$.

In addition, a marginal interaction of Form, Meaning, Hemisphere, and Region, $F(2,46) = 2.90$, $p = .07$, in the overall analysis suggested a specific topography of early meaning-related effects. A three-way ANOVA including Form, Meaning, and Region revealed a significant interaction of all three factors at right-hemisphere leads, $F(2,46) = 4.35$, $p = .04$. Only contrast analysis for medial right electrode leads confirmed that unrelated words (iv) differed from the remaining conditions (i)–(iii), Wilks' lambda adjusted $F(3,23) = 6.53$, $p = .02$. As demonstrated by post hoc comparisons, unrelated words (iv) elicited more positive ERPs than the remaining conditions for right medial electrode leads: (i) completely matching words, $t(23) = 4.12$, $p = .05$; (ii) meaning-matching words, $t(23) = 4.54$, $p = .05$; and (iii) form-matching words, $t(23) = 7.92$, $p = .01$. Crucially, Model (c) in which completely matching words (i) were tested against the other conditions (ii)–(iv) was not confirmed by contrast analyses for any region of interest (ROI), all Wilks' lambda adjusted $F(3,23) \leq 2.09$, $p \geq .16$. Furthermore, Conditions (i), (ii), and (iii) did not differ from each other in any ROI, all $t(23) < 1$, suggesting that form and meaning information do not cumulate in the early right-hemispheric ERP effect. Taken together, results for the P220 effect confirm assumption (B) (form-and-meaning-based activation).

Subtracting each type of match—Conditions (i), (ii) and (iii)—from mismatch (iv) revealed characteristic positive-going difference waves starting at 200 msec and peaking at 220 msec over the right hemisphere with a maximum over temporocentral leads, henceforth referred to as the P220 difference wave (see Figure 3).

Integration of Form and Meaning

In the time window between 300 and 600 msec we report significant effects for the factor Meaning, $F(1,23) = 31.22$, $p < .001$, and for interactions of the factors Form and Meaning, $F(1,23) = 15.42$, $p < .001$; Form and

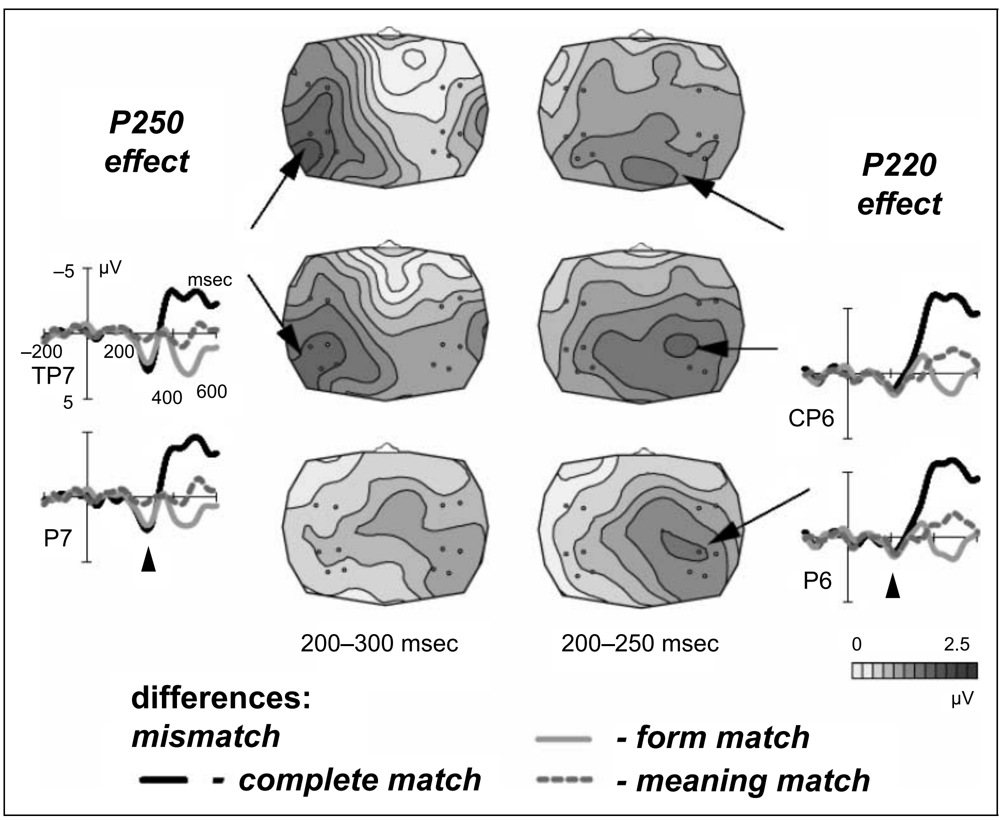

Figure 3. Neurophysiological correlates of form-dependent activation (P250 effect) and form-and-context-dependent activation (P220 effect). ERPs for different types of matching words were subtracted from ERPs for mismatch: mismatch – complete match (black solid line); mismatch – form match (gray solid line); mismatch – meaning match (gray dashed line). Spline interpolated isovoltage maps of the difference waveforms illustrate scalp distributions of the P250 effect (left) and of the P220 effect (right).

Region, $F(2,46) = 6.19$, $p = .01$; Meaning and Region, $F(2,46) = 7.19$, $p = .01$; and Form, Meaning and Region, $F(2,46) = 16.81$, $p < .001$. Separate two-way ANOVAs for each ROI revealed significant interactions of Form and Meaning in all regions: frontal, $F(1,23) = 6.33$, $p = .02$; temporocentral, $F(1,23) = 15.95$, $p < .001$; and parieto-occipital, $F(1,23) = 11.32$, $p = .01$. Contrast analyses showed that completely matching words (i) differed from the other conditions (i)–(iii) in all ROIs: frontal, Wilks' lambda adjusted $F(3,23) = 15.44$, $p < .001$; temporoparietal, Wilks' lambda adjusted $F(3,23) = 18.01$, $p < .001$; and parietooccipital, Wilks' lambda adjusted $F(3,23) = 12.04$, $p < .001$. Post hoc analyses demonstrated that with the exception of the meaning match condition (ii) over the anterior region, all conditions (ii) to (iv) elicited a significantly enhanced negativity as compared to completely matching words (i), all $t(23) \geq 11.46$, all $p \leq .001$. Taken together, N400 amplitude reflects account (D) (integration of form-and-meaning).

DISCUSSION

The present study investigated temporal activation patterns related to bottom–up and top–down flow of information in speech recognition. By means of visual target words we tested activation of lexical representations that (i) completely matched a preceding spoken sentence with an incomplete final word; (ii) matched the meaning of the sentence, but not the incomplete word; (iii) matched the incomplete word, but not the sentence context; or (iv) were completely unrelated to a preceding sentence fragment. ERPs recorded to target words reflect snapshots of temporal activation patterns in online speech recognition. Two early ERP effects can be distinguished, which we will relate to the simultaneous activation of two different groups of lexical representations. A left-hemispheric P250 difference wave reflects bottom–up activation. A right-hemispheric P220 difference wave reflects both bottom–up and top–down activation. In addition, we observed that the N400 amplitude is modulated by the integration of form and meaning information.

Cohort Activation

Assumptions tested in the present experiment critically depend on the detection of neurophysiological correlates of cohort activation in sentence fragment priming. We have previously shown that visual probes immediately following spoken word fragments elicit different ERPs when they belong to a cohort activated by the fragment than when they do not belong to the activated cohort (Friedrich, Kotz, Friederici, & Gunter, 2004). The so-called P350 effect, a left-hemispheric difference potential ranging from 200 to 400 msec, varies as a function of goodness of fit between the signal and the lexical representation (Friedrich, 2005; Friedrich, Kotz, Friederici, & Alter, 2004).

Comparable to the P350 effect in word fragment priming, we also observed a left-hemispheric difference potential in sentence fragment priming. The so-called P250 effect differentiated form-matching and mismatching words. Based on the sensitivity of the P250 to form overlap, and similar onset and scalp distribution of P350 and P250 difference waves, we relate the P250 effect to the activation of multiple form-matching candidates. The earlier peak and offset of the P250 might be a consequence of an earlier onset of incremental processes in sentence recognition. This is confirmed by an early onset of the N400 effect in the present study.

Although we report evidence for purely form-based neuronal activation, no ERP effect could be related to purely meaning-based neuronal activation. Nevertheless, in addition to the P250 effect we found a larger positivity for unrelated words than for form-matching, meaning-matching, and form-and-meaning-matching words in a right-hemispheric P220 effect. Based on its simultaneous onset with the P250 we assume that the P220 is also related to lexical activation and is specific to sentence fragment priming. Together P250 and P220 effects indicate that more than one cohort is activated during word identification in spoken sentences. Left-hemispheric P250 difference waves predominantly reflect bottom–up activation of tokens that are consistent with the speech signal of an uttered word. This type of neural activation supports assumptions of modular models of speech recognition in which information flows from bottom to top levels (Norris et al., 2000; Marslen-Wilson, 1987, 1989). Right-hemispheric P220 difference waves are equivalently related to both bottom–up activation of tokens that are consistent with the speech signal and top–down activation of tokens that are consistent with the ongoing sentence interpretation. Although top–down and bottom–up activation start simultaneously, purely form-based activation lasts longer.

Taken together, the P250 and the P220 effects indicate that multiple tokens of a single representation are activated in parallel by different neural populations. Different cohorts can be separated on the basis of their topography, their sensitivity to top–down activation, and their time courses. The assumption of parallel cohort activation integrates into an emerging view of multiple parallel processing streams in the auditory system that has been specified for speech perception (Scott, 2005). Furthermore, the notion that multiple tokens of a particular lexical entry are activated during the time course of lexical access has been explicitly formulated in the Shortlist model of speech recognition (Norris, 1994). The present results additionally suggest that simultaneously activated tokens are differentially sensitive to bottom–up and top–down mechanisms.

One may speculate about a direct relationship between the observed lateralization of P250 and P220 effects and the lateralization of underlying neural generators. This suggests that initial cohort activation is guided by word form information in both hemispheres, but by incremental sentence interpretation only in the right hemisphere. A specific role of the right hemisphere in meaning recognition is supported by patient studies. It has been shown that right-hemisphere-damaged patients exhibit subtle deficits in comprehending the relationship between an utterance and its context (Joanette, Goulet, & Hannequin, 1990), or that the ability to use sentence context during lexical integration is preserved in patients with left-hemisphere damage (Gindrod & Baum, 2005; Schwaab, Brown, & Hagoort, 1997). Evidence for different strategies of meaning use in both hemispheres also originated from behavioral visual half-field studies (see Faust, 1998, for a review). By applying this technique to ERP research, subtle hemifield differences in the brain response to several aspects of sentence comprehension, such as detection of lexical associations and category structure, or processing of jokes, have been shown (Coulson, Federmeier, Van Petten, & Kutas, 2005; Coulson & Williams, 2005; Federmeier & Kutas, 1999). Detailed source localization that allows one to specify whether both hemispheres really activate and maintain different sets of lexical tokens awaits future research.

Integrative Processes

The present ERP results suggest that the integration of form and meaning does not occur during initial cohort activation, as the P220 effect for completely matching words did not show additional enhancement relative to the other conditions. Apart from its early activation, integration of form and meaning was reflected in the N400 amplitude. Obviously, a combination of both form and meaning match effectively reduces the N400 amplitude in sentence fragment priming. That is, cumulative effects, as proposed by interactive or distributed models of speech recognition (Gaskell & Marlsen-Wilson, 1997; McClelland & Elman, 1986), appear to occur relatively late.

According to the interpretation of the N400 (see Kutas & Federmeier, 2000), pure meaning match or pure form match requires more processing effort than does a complete match. This finding is paralleled by ERP studies on sentence comprehension with sentence final words that share their onset with expected words, for example, *It was a pleasant surprise to find that the car repair bill was only seventeen dolphins* (Van Petten et al., 1999; see Introduction). Following an initial phase of reduced negativity, N400 amplitude rises as soon as *dolphins* no longer matches the expected word *dollars* (van den Brink & Hagoort, 2004; van den Brink et al., 2001). Together, the present ERP results on sentence fragment priming and the previous results on sentence comprehension show that tokens that are spuriously activated either by the speech signal or by the context, but which are not selected for further processing, are difficult to integrate. This result may be related to competition between activated representations, an assumption that is implemented in most models of spoken word recognition (Gaskell & Marslen-Wilson, 1997; Norris, 1994; Luce, Piosini, & Goldinger, 1990; McClelland & Elman, 1986). The final selection of completely matching representations may inhibit further processing of pure form or pure meaning match.

Implications for Behavioral Data

Our behavioral data provide additional support for the interpretations resulting from the ERP data. Reaction times and error rates indicate that neither form nor meaning information can be ignored in speech recognition. It was difficult for participants to accept context congruent words that differed from preceding word fragments in form. Similarly, participants had difficulty rejecting words that matched the word fragment but not the sentence. Based on the ERP results, we propose that the behavioral data were biased by at least two processes: automatic activation and inhibited integration. Rejection of purely form-matching words was biased by automatic activation of those words, whereas acceptance of purely meaning-matching words was biased by late inhibition of those words. In line with this, acceptance of completely matching words profited from both initial activation and facilitated integration, whereas rejection of unrelated words was neither biased by activation nor biased by facilitated integration.

The present ERP results may provide advice for future research on speech recognition. Crucially, the assumption of parallel activation of separate cohorts during speech recognition appears to be difficult to disentangle from methods that record categorical responses. More specifically, reaction times for lexical decisions in semantic priming tasks, which support purely form-based lexical activation (Connine, Blasko, & Wang, 1994; Zwitserlood, 1989; Onifer & Swinney, 1981; Swinney, 1979; Tanenhaus et al., 1979) and steady eye fixations of one object or another, which confirm immediate incremental processes during lexical activation (Dahan & Tanenhaus, 2004), might be biased by different temporary activation patterns in the mental lexicon. As both techniques apparently do not allow one to track parallel activation patterns during word identification, we suggest that the assumption of functional parallelism in spoken word recognition needs to be investigated with multidimensional online measures.

Conclusion

Using ERPs recorded during sentence fragment priming, we show that a form-based set of lexical tokens and a form-and-meaning-based set of lexical tokens are simultaneously made available to higher order processes of speech recognition. The involved processing streams can be separated at a neural level by means of scalp distribution and time course of activation. Bottom–up information appears to be shared by two lexical networks that are activated during early word identification. Top–down information appears to be coded by only one of both activation networks. Our results indicate that modeling of human speech recognition has to integrate both types of information flow in parallel activated cohorts. In addition, it must implement the integration of both form and meaning as a function of processes that follow initially parallel activation of multiple tokens in several cohorts.

Reprint requests should be sent to Claudia Friedrich, Biological Psychology and Neuropsychology, University of Hamburg, Von-Melle-Park 11, D-20146 Hamburg, Germany, or via e-mail: Claudia.Friedrich@uni-hamburg.de.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the rime course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Bowers, J. S., & Davis, C. J. (2004). Is speech perception modular or interactive? *Trends in Cognitive Sciences, 8,* 3–5.

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1988). Modified nomenclature for the "10%" electrode system. *Journal of Clinical Neurophysiology, 5,* 183–186.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics, 56,* 624–636.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potentials reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6,* 624–636.

Connolly, J. F., Phillips, N. A., Steward, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43,* 1–18.

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 129–147.

Coulson, S., & Williams, R. W. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia, 43,* 128–141.

Cutler, A., & Clifton, C., Jr. (2000). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–155). Oxford: Oxford University Press.

Cutler, A., & van Donselaar, W. A. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44,* 171–195.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513.

Fodor, J. (1983). *Modularity of mind.* Cambridge: MIT Press.

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

Faust, M. (1998). Obtaining evidence of language comprehension form sentence priming. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (pp. 161–186). Hillsdale, NJ: Erlbaum.

Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research, 8,* 373–392.

Friedrich, C. K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neuroscience, 6,* 64.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cognitive Brain Research, 20,* 300–308.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience, 16,* 541–552.

Gaskell, M. G., & Marlsen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gindrod, C. M., & Baum, S. R. (2005). Hemispheric contributions to lexical ambiguity resolution in a discourse context: Evidence form individuals with unilateral left and right hemisphere lesions. *Brain and Cognition, 57,* 70–83.

Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review, 107,* 735–767.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia, 38,* 1518–1530.

Joanette, Y., Goulet, P., & Hannequin, D. (1990). *Right hemisphere and verbal communication.* New York: Springer-Verlag.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

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

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122–147). Cambridge: MIT Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71–102.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sounds onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge: MIT Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3–36). Hillsdale, NJ: Erlbaum.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.
Norris, D., McQueen, J. M., & Cutler, D. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299–370.
Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency-of-meaning and contextual bias. *Journal of Verbal Learning and Behavior, 17,* 225–236.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.
Pylkkänen, L., Linás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience, 18,* 97–109.
Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7,* 187–189.
Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81,* 666–678.
Schwaab, T. Y., Brown, C., & Hagoort, P. (1997). Spoken sentence comprehension in aphasia: Event-related brain potential evidence for integration deficit. *Journal of Cognitive Neuroscience, 9,* 39–66.
Scott, S. K. (2005). Auditory processing—Speech, space and auditory objects. *Current Opinion in Neurobiology, 15,* 197–201.
Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412–432.
Spivey, M. J., Grosejean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 10393–10398.
Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Memory and Behavior, 18,* 645–659.
Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32,* 359–372.
Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence of multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Memory and Behavior, 18,* 427–440.
van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–985.
van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16,* 1068–1084.
Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394–417.
Zwitserlood, P. (1989). The locus of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25–64.
Zwisterlood, P. (1999). Spoken words in sentence contexts. In A. D. Friederici (Ed.), *Language comprehension: A biological perspective* (pp. 71–99). Berlin: Springer.