Journal of Experimental Psychology:
Learning, Memory, and Cognition
1999, Vol. 25, No. 2, 394–417

0278-7393/99/$3.00

Time Course of Word Identification and Semantic Integration in Spoken Language

Cyma Van Petten, Seana Coulson, Susan Rubin, Elena Plante, and Marjorie Parks
University of Arizona

The minimum duration signal necessary to identify a set of spoken words was established by the gating technique; most words could be identified before their acoustic offset. Gated words were used as congruous and incongruous sentence completions, and isolation points established in the gating experiment were compared with the time course of semantic integration evident in event-related brain potentials. Differential N400 responses to contextually appropriate and inappropriate words were observed about 200 ms before the isolation point. Semantic processing was evident before the acoustic signal was sufficient to identify the words uniquely. Results indicate that semantic integration can begin to operate with only partial, incomplete information about word identity. Influences of semantic constraint, word frequency, and rate of presentation are described.

Human speech proceeds at a rate of some 15 phonemes, or more than three words per second. The acoustic signal specifying each phoneme varies with the surrounding phonemes, the speaker, and the rate and loudness of speech. Few, if any, silent periods mark the boundaries between successive words. How listeners solve the perceptual problem of identifying words in fluent speech continues to preoccupy researchers. However, grasping the meaning of a single sentence requires not only segmenting the acoustic stream into words but also accessing the meaning of each word, picking the relevant aspect of that meaning for the present context, evaluating the grammatical relationships among words, and determining if the speaker's tone of voice indicates sarcasm, humor, or sincerity. The speed and continuous nature of the acoustic input suggests that there must be some temporal overlap among these diverse processes, but there has been considerable dissent among psycholinguists about the relationship between word identification and semantic integration.

At one extreme are theories that specify that a word must securely be identified from sensory input alone, and its context-free meaning accessed, before that meaning can be made available for integration with prior semantic context (Bradley & Forster, 1987; Garrett, 1990). This viewpoint has been based primarily on studies conducted with printed words (Forster, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Seidenberg, Waters, Sanders, & Langer, 1984). In contrast, some accounts of spoken word processing allow semantic processing to be initiated as perceptual processing of the input continues. We do not attempt an exhaustive review of models of spoken word recognition here but instead point to only a few contrasting ideas and empirical results relevant to the present study (see Altmann, 1990; Frauenfelder & Tyler, 1987; Lively, Pisoni, & Goldinger, 1994; Massaro, 1994, for more extensive reviews).

One general idea expressed in several theories is that a set of lexical candidates is activated by incomplete auditory input. Hearing *CAP*—will thus activate a set of acoustically consistent words such as CAPTAIN, CAPTIVE, CAPTION, CAPITAL, and CAPSULE (Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). Many theorists have addressed the consequences of such multiple activation for recognition of the true word and have taken different positions with regard to competitive interactions among active candidates (see McQueen, Cutler, Briscoe, & Norris, 1995, for a recent review). The focus of the present study was the semantic nature of these early activations. Questions remain as to the semantic status of these incompletely specified word candidates: Are their meanings "activated" in some sense, and, if so, do these meanings make contact with the overall meaning of a sentence or discourse context?

Several experimental techniques have been brought to bear on these questions. In the gating technique, participants are presented with only the first 50 ms of a word and are

Cyma Van Petten, Department of Psychology and National Center for Neurogenic Communication Disorders, University of Arizona; Seana Coulson, Susan Rubin, and Marjorie Parks, Department of Psychology, University of Arizona; Elena Plante, Department of Speech and Hearing Science and National Center for Neurogenic Communication Disorders, University of Arizona.

A preliminary report was presented at the 6th International Conference on Cognitive Neuroscience, Asilomar, California, June 1996. This work was supported by Grant NS 30825 from the National Institute for Neural Disorders and Stroke; by grants to the University of Arizona Center for Cognitive Neuroscience from the McDonnell Foundation, Pew Charitable Trust, and the Flinn Foundation; and by Grant DC 00077 from the National Institute for Deafness and Communication Disorders. We are grateful to Ava Senkfor, Ron Ohst, Mark Bakarich, and Jonathan Forster for technical support and to Jonathan Folstein and Kathleen Guess for assistance in stimulus preparation.

Correspondence concerning this article should be addressed to Cyma Van Petten, Department of Psychology, University of Arizona, Tucson, Arizona 85721. Electronic mail may be sent to vanpettc@u.arizona.edu.

asked to guess what it might be, or the first 100 or 150 ms, and so forth until the end of the word. With brief stimuli, many candidate words are produced, but listeners converge on the correct word, given sufficient acoustic input (Grosjean, 1980; Tyler & Wessels, 1985). The *gate,* or word fragment duration at which most participants produce the correct word, is known as the *isolation point.*[1] When the auditory fragment is preceded by a sentence context, less acoustic input is required to produce the target word, and fewer contextually inappropriate responses are generated at early gates than if the context imposes no semantic constraints (Craig, Kim, Rhyner, & Chirillo, 1993; Grosjean, 1980; McAllister, 1988; Salasoo & Pisoni, 1985; Tyler & Wessels, 1983; Zwitserlood, 1989). This result may be used to argue that semantic context is combined with incomplete sensory information to narrow in on the target word (Tyler & Wessels, 1983). However, the data do not mandate this conclusion. After all, participants can generate contextually appropriate sentence completions with no sensory information in a cloze task in which the final word is replaced with a blank or silence. A similar argument has been made for a technique which is similar to gating in that listeners are asked to identify words, given perceptually ambiguous stimuli. Miller and Connine have both asked listeners to identify items in which the initial consonants were altered to be midway between voiced and voiceless (e.g., *B/PATH*). When a preceding context is consistent with one of the alternatives (e.g., "She needed hot water for the ___"), listeners favor the contextually consistent possibility (Connine, 1987; Connine, Blasko, & Wang, 1994; J. L. Miller, Green, & Schermer, 1984). However, Connine has argued that this reflects a postperceptual bias given that the nonlinguistic manipulation of monetary payoff ratios serves to alter participants' responses in much the same way.

A rather different method for examining the semantic status of incomplete word candidates is to determine whether they are capable of "priming" subsequent words. Zwitserlood (1989) looked for downstream influences of early semantic processing by presenting fragments like GEN— after a sentence frame that could plausibly be terminated by GENEROUS but not GENERAL. The dependent measure was lexical-decision time for visual probe words such as GIFT and ARMY.[2] When the auditory fragment was too short to differentiate GENEROUS from GENERAL in a gating study, both related probe words yielded faster reaction times (RTs) than a control condition. Only when the auditory fragment duration was equal to the IP established in the prior gating experiment did the sententially relevant probe GIFT gain an advantage over the irrelevant probe ARMY. Because the brief auditory fragments provided an effective semantic context for the probe items, the first result supports the view that semantic processing begins with incomplete sensory information. However, the second result suggests that this early semantic processing was not constrained by the sentence context, so that semantic integration is postponed until the acoustic signal is sufficient to identify a word uniquely.

The temporal separation between context-free lexical semantics and sentential integration observed in Zwitserlood's (1989) study is similar to the conclusions of experiments investigating ambiguity resolution in the visual modality (reviewed by Simpson, 1994; see also Sereno, 1995; Van Petten & Kutas, 1987). However, two additional studies make these conclusions less secure. Chwilla (1996) also used a cross-modal presentation technique, with auditory word fragments that were truncated either before or after the context-free recognition point established without context. She found no RT advantage for related visual words after short auditory primes; context effects were observed only when the auditory signal was long enough to specify the word. Although this experiment did not include sentence contexts, it calls into question the notion of any early semantic processing. Moss and Marslen-Wilson (1993) used a paradigm much like Zwiterslood's, which consisted of a sentence frame, an auditory prime word, and a visual target, but varied the nature of the semantic relationship between the prime and the target. When the target had a strong associative relationship to the auditory prime (e.g., CHICKEN–HEN), lexical-decision times were faster than for unrelated words. This RT advantage was observed after auditory fragments that were some 130 ms shorter than the recognition point measured for these words in the absence of context. The semantic priming effect from incomplete auditory primes held even when the preceding sentence frame did not provide any contextual support for the auditory fragment (e.g., "When she was looking through the photographs, Tracey found a rather odd one of some chi___"). This result replicates Zwitserlood's in suggesting that semantic activation of a spoken word begins before the acoustic input can uniquely specify which word it is, but that this early semantic processing involves lexical semantics only, rather than integrative processing. However, target words that were less strongly related to the auditory prime produced a different pattern of results. In this case, the RT advantage produced by short auditory primes was restricted to targets that matched the semantic properties suggested by the sentence context. For example, only FARM showed a priming effect after a sentence concerning places where chickens might live, but only BEAK showed a priming effect after a sentence about chickens catching worms. This aspect of Moss and Marslen-Wilson's results suggests that the meaning of the auditory fragment was not context free but had already begun to be integrated with the preceding sentence.

Taken together, the results of these three studies are

[1] A related measure, *recognition point,* incorporates participants' confidence ratings and the accuracy of their responses. The recognition point, when participants produce highly confident and accurate responses, typically requires somewhat longer acoustic inputs than the IP. It is not clear if the difference between the two measures can be taken as a constant or is variable among listeners and stimulus sets. In the present study, we used the IP measure to evaluate the briefest acoustic input that leads to accurate identification, and this measure will be referred to in the discussion of other studies whenever available.

[2] Both Zwitserlood's (1989) and Chwilla's (1996) experiments were conducted in Dutch, but the English examples illustrate the relevant contrasts.

compatible with any position one might wish to take: (a) Semantic processing of spoken words begins only when the acoustic signal is capable of picking out one word; (b) semantic properties of multiple candidates are activated before the sensory input can uniquely identify one word, but no attempt is made to integrate these properties with the ongoing sentence; or (c) semantic processing begins early in a spoken word, and so does sentential integration (Chwilla, 1996; Moss & Marslen-Wilson, 1993; Zwitserlood, 1989).

The studies reviewed above illustrate the two most common classes of experimental techniques used to investigate the time course of semantic processing in spoken language. One class of techniques seeks to answer the question of whether semantic processes can exert an influence on incomplete perceptual analyses. When such evidence exists, it may support the claim of early semantic processing. The production of semantically appropriate words in a gating task in which the acoustic input alone is insufficient is one example of this sort of evidence (Tyler & Wessels, 1983). The report of an ambiguous phoneme as forming a semantically congruous word is another example (Connine, 1987; J. L. Miller et al., 1984). The same logic underlies the invention of other clever experimental measures, such as the correction of mispronunciations in shadowing coherent prose (Marslen-Wilson & Welsh, 1978), and failures in noticing that phonemes have been replaced by noise when the carrier word is semantically supported (Samuel, 1981). However, evidence of this sort may also stimulate arguments as to whether the dependent measure indeed evaluates perceptual processes or is subject to bias from other sources, arguments that have raged for each of the measures above (Connine, 1990; Tyler, 1990). It may be useful to answer a logically prior question—When does semantic processing begin?—independent of its possible consequences for perceptual analyses.

The second general class of methodologies relies on semantic-context effects to reveal when spoken words begin to make contact with semantic representations. The logic here is that if an incomplete auditory word yields an RT advantage for subsequent related words, it must itself have been processed at a semantic level. This inference is unassailable but does not yield unambiguous data about when the semantic processing occurred. Over the past 3 decades, a number of possible mechanisms for semantic-context effects have been proposed. Some of these incorporate true priming effects, in that the presentation of the context item raises the activation level of the target item before the target is actually presented (Collins & Loftus, 1975; Morton, 1969). If context effects from brief auditory primes were due to this sort of mechanism, it would indeed imply that the incomplete prime activated a full word (or set of words) together with targets related to that word. However, other proposed mechanisms place context effects in the interaction of prime and target after both have been presented (Neely & Keefe, 1989; Norris, 1986; Ratcliff & McKoon, 1988). If context effects occur by one of these latter mechanisms, the ability of incomplete auditory words to provide effective context can lead to only two secure conclusions: (a) the auditory context was processed semantically, and (b) the semantic processing occurred sometime during the interval between its presentation and the participant's response. The latter conclusion places some constraints on time course, but an upper limit set by the participant's RT is not as useful as one would like, given the apparent discrepancy between rapid comprehension and slower manual or oral RTs.

Some investigators have thus turned to measures that promise greater immediacy and the ability to place a lower bound on comprehension time. Eye movements are typically somewhat faster than either manual responses or pronunciation times. One recent study indicates that listeners can make an eye movement to the referent of a noun as quickly as 145 ms after the offset of the spoken word (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). But even this rapid response is likely to overestimate comprehension time, given some delay in programming the motor response.

The Present Study

In the present study, we investigated the time course of semantic processing by an event-related potential (ERP) measure of brain activity that can be elicited well in advance of a behavioral response and that can yield high temporal resolution. In Experiment 1, we established the IP for a set of 704 words. In Experiment 2, ERPs were recorded when these words served as congruous or incongruous completions of spoken sentences. The IP measure was used to establish the minimum duration of acoustic information necessary to identify a word in the absence of semantic context. By comparing this measure with the latency of the electrophysiological sentence congruity effect, we hope to provide new information about the temporal relationship between word identification and semantic integration in speech comprehension.

ERPs are summed synaptic potentials recorded from the scalp, which are extracted from the background electroencephalogram by averaging responses time locked (typically) to stimulus onset (Regan, 1989). Previous research has shown that printed, spoken, and signed words elicit an ERP component which is negative in polarity and typically peaks at about 400 ms after stimulus onset (N400). The N400 is smaller in amplitude when words are preceded by a congruent sentence frame, or a related single word, than if the eliciting word is incongruous or unrelated to the preceding context (Kutas & Hillyard, 1980a, 1980b, 1980c; see Kutas & Van Petten, 1988, 1994, for reviews). With sentence contexts, the largest possible differences in N400 amplitude are observed in comparisons between completely predictable words and wholly anomalous completions. It is important to note, however, that N400 amplitude does not index anomaly but absence of contextual support. For congruous sentence terminal words, N400 amplitude follows a graded function which is closely (and negatively) correlated with cloze probability (Kutas & Hillyard, 1984). Within congruous sentences isolated from discourse context, words that occur in late-sentence positions elicit smaller N400s than earlier words because they can benefit from a larger chunk of the ongoing sentence context (Van Petten, 1995; Van Petten

& Kutas, 1990). In contrast, the amplitude of the N400 elicited in syntactically legal but semantically incoherent sentences is uniformly large and does not vary as a function of position in the sentence (Van Petten, 1993; Van Petten & Kutas, 1991; Van Petten, Weckerly, McIsaac, & Kutas, 1997).

In the present study, the sensitivity of the N400 to semantic context was used to establish the temporal relationship between the moment a word can uniquely be identified and the initiation of semantic integration. If semantic integration begins only after word identification is complete, then responses to words that fit a sentence context should begin to differ from words that do not fit shortly after the IP. Alternatively, if semantic processes begin to operate on the partial and incomplete results of perceptual analyses, then the influence of semantic context might begin before the IP, as soon as the acoustic input becomes discrepant with semantic expectations. The experimental strategy applied in our experiments differs from the experiments reviewed earlier in two regards. First, we did not attempt to evaluate whether semantic processing alters ongoing perceptual analyses. It may be possible to address such a question given secure knowledge of the time frame of the two sets of processes, but we attempted to establish their relative time courses first. Second, the current strategy relies on the ubiquitous semantic-context effect observed in many dependent measures. However, unlike the cross-modal priming studies described earlier, our focus was on the auditory word itself, rather than its influence on subsequent items.

In Experiment 1, we describe the gating technique used to establish word identification times, and we evaluate the degree to which the IP measure reflects an all-or-none threshold for identification. IPs were measured in the absence of semantic context to quantify the duration of the sensory signal necessary for word identification. In Experiments 2 and 3, ERPs were recorded to the same words when placed in sentence contexts.

Experiment 1

The central comparison of the present study was between word identification times as indexed by the isolation point in gating and semantic integration times as indexed by ERP latencies. The primary purpose of Experiment 1 was thus to establish IPs for the words to be used in Experiment 2. However, this required gating some 700 words, a substantially larger number than have been used in prior research (typically less than 30, never more than 100). The larger stimulus set allowed both extension and replication of prior results and some more detailed analyses that were less amenable to smaller databases.

One general issue was central to the design of the overall study—the psychological significance of the IP measure. The term *point* implies a discrete moment in time where a threshold is crossed. However, in empirical studies, IPs are necessarily defined by an arbitrary accuracy criterion when 70% to 90% of the participants report the target word. Does such a criterion reflect a true discontinuity in listeners' ability to identify the target word, or is the threshold an artifact of the measurement? In fact, the results of previous gating studies display smooth accuracy functions wherein likelihood of reporting the target word increases monotonically with stimulus duration up to some asymptote (Tyler & Wessels, 1983; Wayland, Wingfield, & Goodglass, 1989; Zwitserlood, 1989). However, these accuracy functions have been collapsed across words with different IPs, which may blur sharp increases in accuracy for individual items at their IPs. The IP is only one of various experimental measures that may evaluate the time course of word identification, but the question is a more basic one. Some aspects of spoken word processing—recognition of phonetic features, phonemes, and prosodic contours—must occur continuously (Stirling & Wales, 1996; Warren & Marslen-Wilson, 1987). Is word identification similarly an incremental process wherein each bit of acoustic information is no more or less useful than the last? Or is there indeed a "magic moment" when the listener can settle on a single word, and all remaining alternatives are dismissed? In the gating experiment, we examined several measures that may indicate a true discontinuity at the IP. The continuous versus discontinuous aspects of word identification evident in the gating results were then contrasted with the measures of semantic integration examined in Experiment 2.

Method

Participants. Two hundred twenty young adults participated for course credit in an introductory psychology class at the University of Arizona. They ranged in age from 18 to 26 years. All were fluent speakers of English and reported no history of neurological disorder or learning disability.

Stimuli. One hundred and seventy-six quartets of words were selected such that two members shared initial phonemes (*cohort* words such as DOLLAR and DOLPHIN), and each of these was matched with a word that shared final phonemes (*rhyming* words such as SCHOLAR and MUFFIN). The two cohort words in each quartet shared, on average, 3.1 phonemes (range, 2–5); the rhyming words shared an average of 2.8 final phonemes with their associated cohort words (range, 2–5). For most purposes of this first experiment, the stimuli can be considered an undifferentiated set of 704 words, but the quartet structure was critical for the design of Experiments 2 and 3, which use the information gathered in this gating experiment. Within each pair of cohort words, one was considered "high" frequency and one "low" frequency, depending on the relative frequency of usage of the two words (Francis & Kučera, 1982). Thus, DOLLAR (frequency of 144) is the high-frequency member of the DOLLAR–DOLPHIN pair, but CHARITY (frequency of 12) is also the high-frequency member of its pair CHARITY–CHARIOT. Mean frequency of usage for the rhyming words was matched to the cohort words, as summarized in Table 1. The number of syllables was the same for each word in a quartet, ranging from one to four across quartets ($M = 1.95$).

Words were initially recorded on analog tape in a male voice and were then digitized with an 11-kHz sampling rate. Each digital word file was then edited to ensure good synchronization between acoustic onset and offset and the beginning and ending of the file and to equate amplitude levels across words. Durations of the stimulus classes are summarized in Table 1. In the second phase of editing, these full-duration versions of each word were transformed into gated versions using 50-ms intervals. The first gate included only the first 50 ms, the second gate 100 ms, and so on until the end of the word. Amplitude over the final 10 ms of each gate file was tapered to zero to avoid acoustic transients (clicks) that would be created by a sharp cutoff.

Procedure. The 8,540 gated stimulus files were randomly arranged into 21 sets of 400 and into 1 shorter set of 140 files. A single trial consisted of the carrier phrase "The next word is," 400 ms of silence, and one gate file. These were recorded on analog tapes, with an intertrial interval of 6 s. Separate groups of 10 individuals listened to each tape and recorded their best guess as to the identity of each gated word on a response form.

Results

Examination of the responses suggested that listeners were often unable to decode the acoustic information offered at the two shortest gates. With 50- and 100-ms signals, 22% of the responses reflected inaccurate identification of the word-initial phoneme. Accuracy of initial-phoneme identification was 89% by the 150-ms gate. At this third gate, the elicited responses shared an average of 1.92 initial phonemes with the actual word.

The isolation point. The IP for each word was defined as the gate at which at least 7 of 10 participants correctly identified the word, and accuracy remained above 60% at subsequent gates. The IP occurred before the end of the word for 96% of the items; on average, it preceded the end of the word by 225 ms. Figure 1 shows the distributions of word durations and IPs across the stimulus set.

Like all psychophysical thresholds, the 70% accuracy criterion used here is arbitrary. Previous gating studies have used two presentation formats. In one format, the same item is presented to each listener in longer and longer versions, and the IP is defined as the shortest stimulus duration that first yields correct identification, averaged across multiple listeners. With a small number of participants, this procedure is vulnerable to underestimation of the "true" IP if one or more listeners fortuitously guess the correct word from inadequate input, or a similar overestimation based on outlying participants who were inattentive at the critical gate. The second format uses a between-subjects design more similar to that used here—presentation of each gate to different listeners; the IP is then defined as the duration when a majority of the listeners correctly identify the word. Our 70% definition of "majority" is in general accord with that used in similar studies (Brown, 1990: 65%–75% of participants; Grosjean, 1980: 5 of 8 participants). Later, we consider how well this definition of the IP reflects the duration of acoustic signal necessary for accurate word identification.

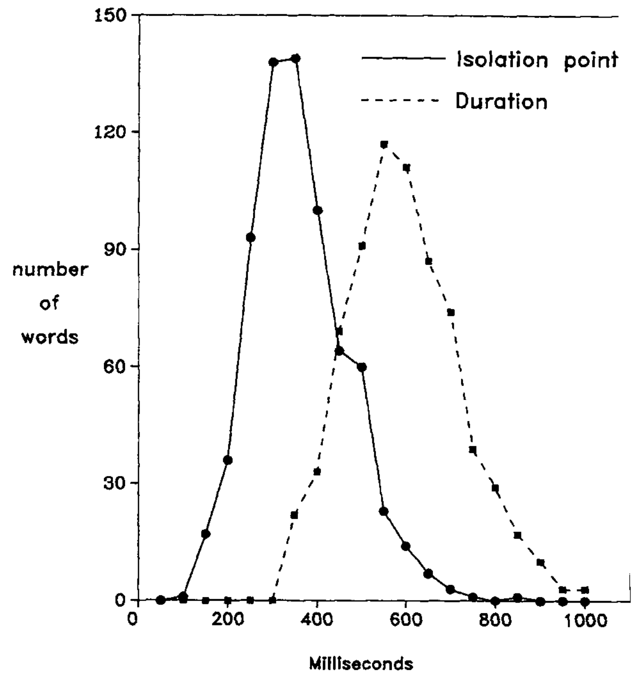

Figure 1. Distributions of full-word duration and isolation points across the set of words, in 50-ms increments (duration was rounded to the closest multiple of 50).

Table 1
Summary of Stimulus Characteristics

Stimuli	Frequency *M*	Frequency *SE*	Duration *M*	Duration *SE*	Isolation point *M*	Isolation point *SE*
Cohort						
High frequency	123	11	604	8	360	8
Low frequency	23	2	607	10	391	8
All cohort	73	6	605	6	376	6
Rhyme						
High frequency	128	13	568	9	332	8
Low frequency	29	4	591	11	356	9
All rhyme	79	7	579	7	345	6

Note. Means, standard errors, duration, and isolation point are presented in milliseconds. $n = 176$ for the high- and low-frequency categories, and $n = 352$ for all cohort and all rhyme categories. Word frequency was calculated as the sum of all regularly inflected forms according to Francis and Kučera (1982).

The IP criterion of 70% yielded a fairly sharp transition in accuracy; identification rates before the IP were low, whereas accuracies after the IP were uniformly high. Figure 2 shows this transition for those words in which it was possible to examine accuracy for at least three gates before the IP (words with IPs of 200 ms or longer) and three gates subsequent to the IP (words whose durations exceeded the IP by at least 150 ms). For these 630 words of the total stimulus set, accuracy increased 53% at the IP, but only 11% at the previous gate, $F(1, 629) = 490.30, p < .0001$.

A second analysis was performed to evaluate the generality of this sharp increase in accuracy at the IP. The accuracy function for each item was examined for any increase between successive gates that exceeded 40%. Of the total set of 704 items, 589 (84%) showed an increase of this magnitude.[3] For these items, the IP assigned by the 70% accuracy criterion matched the discontinuity point in 71% of

[3] Ninety-seven percent of the items showed accuracy increases of at least 30% between successive gates. However, it is not obvious how large an increase needs to be to count as a sharp discontinuity; we used 40% as a conservative measure.

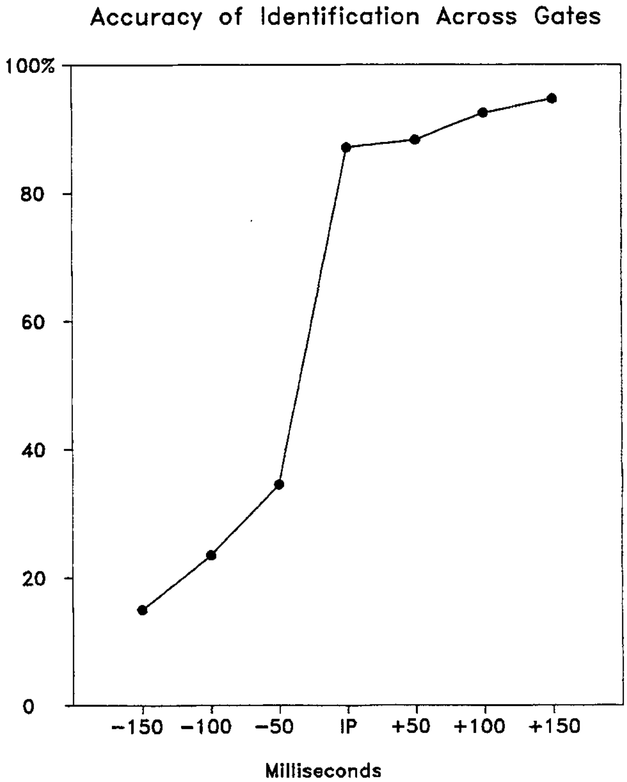

Figure 2. Identification accuracy before, at, and after the isolation point (IP). Accuracies are for 630 words of the total set, with IPs of 200 ms or greater and durations at least 150 ms longer than the isolation point.

the cases. Mismatching cases were those in which accuracy increased sharply across gates but did not reach 70% until one gate after the increase (e.g., for BARLEY, accuracy at the fourth, fifth, and sixth gates was 10%, 60%, and 90%, respectively, so that the largest discontinuity occurred at 250 ms, but the IP was defined as 300 ms). In other terms, the mean discontinuity point was reached at 335 ms, but the assigned IP was 358 ms. Overall, these analyses suggest that the 70% accuracy criterion corresponds fairly well with a sharp increase in ability to identify the target words and only slightly underestimates the time at which this increase occurs.

Cohort size across time. The response and scoring procedures for quantifying the IP provide only partial information about how cohort size changes with increasing acoustic input. The IP measure specifies when a single remaining word is favored by the majority of participants, but the size of the cohort just before the IP may consist of two candidates, three candidates, or a larger number. To evaluate the change in cohort size over time, we counted the total number of candidate words generated at each gate up to the IP (e.g., the *elicited cohort*) for each word, with a final IP between 250 and 450 ms. These items were selected because the sample size for each of these five IPs was large enough to analyze independently ($n = 50$ or more). Given that 10 participants heard each gate, the number of candidates could vary from 1 to 10.

Figure 3 shows the size of the elicited cohort across gates. Before the isolation gate, the decline in cohort size was essentially linear with increasing acoustic information. For words with an IP of 250 ms, the data before the IP were analyzed by an analysis of variance (ANOVA) using item as the random factor and trend analyses to assess changes in cohort size across the four gates from 50 to 200 ms. Similar analyses were used for words with final IPs of 300, 350, 400, and 450 ms. These analyses showed significant linear components for each of the five subsets of words (Fs > 200 for the linear effect of gate duration; Fs < 2 for quadratic and higher order trend components).

Figure 3 also suggests a discontinuity at the IP—a steeper decline in the number of candidates, given the critical 50 ms of input separating the isolation gate from the immediately prior gate. This discontinuity was evaluated by measuring the difference in cohort size for each pair of successive gates. On average, 2.15 words dropped out of the elicited cohort at the isolation gate, compared with 0.84 words at the earlier gates. Post hoc Scheffé tests indicated that this difference in dropout rate was significant ($p < .05$) for words with IPs of 250, 300, 350, and 400 ms, although not for the subset of words with a 450-ms IP.

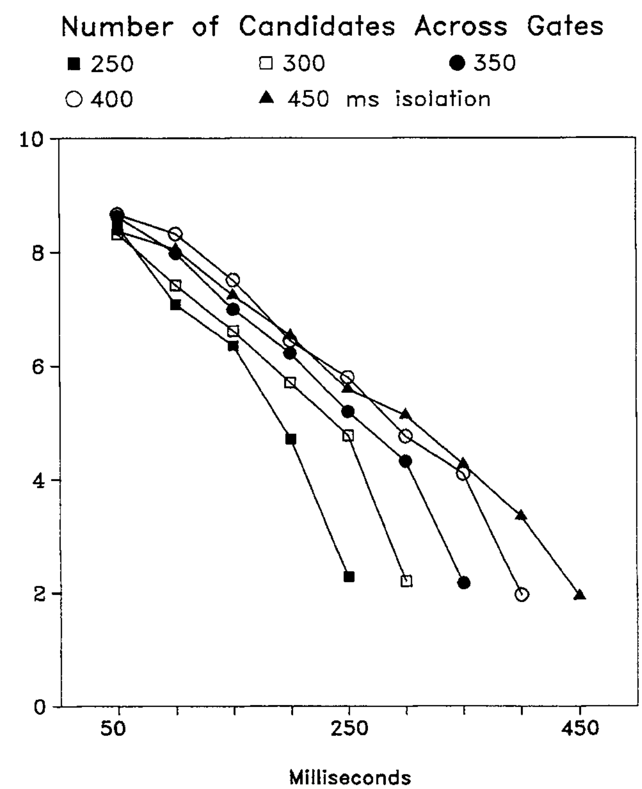

Figure 3. Number of candidate words generated by the participants at gates of increasing length, for subsets of stimuli with different final isolation points. The maximum number of candidates is 10, the number of participants hearing each gated stimulus.

Word frequency and the isolation point. The stimulus set included pairs of words with the same initial phonemes, with each member of the pair categorized as higher or lower in relative frequency (the cohort pairs, see Table 1). On average, the higher frequency members had earlier IPs, by 31 ms (360 vs. 391 ms), $F(1, 175) = 8.74, p < .005$. A more conventional analysis considered frequency independent of the assignment of words to pairs and included all of the stimuli (both the cohort words and their rhyming controls). A frequency cutoff of 30 provided a division into two nearly equal subsets: less than 30, $n = 344$, average frequency of 10.5, $SE = 0.5$; greater than or equal to 30, $n = 360$, average frequency of 139, $SE = 8$. Words with high absolute frequency had an earlier IP than those with low (350 vs. 372 ms), $F(1, 702) = 6.84, p < .01$.

Discussion

The gating results are in general accord with previous results. Given that the words were spoken in isolation with clear word boundaries, the large majority of words could be identified well before their offset (Grosjean, 1980). The minimum signal duration to obtain reliable responses that matched at least the first phoneme of the target word was about 150 ms (Salasoo & Pisoni, 1985; Tyler, 1984). High-frequency words had shorter IPs than low-frequency words (Grosjean, 1980; Marslen-Wilson, 1990).

The most important results of the gating experiment are those that suggest that the IP measure approximates a genuine threshold in word identification. Three findings indicated that the IP measure indexes a discontinuity in the analysis of continuous auditory input. First, the accuracy function had a marked inflection at the IP (Figure 2). Second, the arbitrary criterion of 70% accuracy reflected sharp increases in accuracy for individual items across gates. Finally, the number of candidate words dropped off steeply at the IP (Figure 3). These qualities of the IP measure suggest that it serves as a good measure of the duration of auditory signal necessary for spoken word identification.

Experiment 2

In Experiment 2, the temporal relationship between word identification and semantic integration was evaluated by comparisons between the IP and the sentence congruity effect or difference between congruous and incongruous sentence completions in the ERP. Previous studies in the auditory modality have shown that semantically incongruous words elicit much larger N400 components than congruous words (Connolly, Stewart, & Phillips, 1990; Ford et al., 1996; Holcomb & Neville, 1991; McCallum, Farmer, & Pocock, 1984). The central focus here is on the onset latency of this sentence congruity effect, which will offer at least an upper bound for the initiation of sentence integration.

A priori, the N400 congruity effect may be elicited by word onset, word offset, IP, or *discrepancy point*—when the acoustic input first becomes an implausible sentence completion. The first two alternatives are unlikely ones for a process sensitive to semantic analyses and can be ruled out on the basis of previous studies. For studies in which sentence-terminal words were recorded in isolation and later spliced into the sentence context, typical onset latencies for the N400 congruity effect are about 200 ms after word onset but well before word offset. When final words are initially spoken and recorded together with the sentence frame, onset latencies are substantially earlier. Under these conditions, the pronunciation of the preceding word contains coarticulatory cues about the sentence-terminal word. A divergence between congruous and incongruous final words is thus observed as early as 60 ms after the defined word onset in running speech (Connolly, Phillips, Stewart, & Brake, 1992; Holcomb & Neville, 1991; Woodward, Ford, & Hammett, 1993).

The remaining alternatives—isolation versus discrepancy points—are both consistent with existing data. For instance, O'Rourke and Holcomb (in press) compared pseudowords that diverge earlier or later from possible words in a lexical-decision task. As in previous studies, they observed larger N400s for pseudowords than for real words, but this effect was delayed for items that only became inconsistent with existing words late in the auditory signal (e.g., STAMPION). Connolly and Phillips (1994) conducted a sentence experiment with conditions that bear some similarity to those of the present experiment. When incongruous final words shared initial phonemes with the congruous completion (e.g., "The gambler had a streak of bad LUGGAGE"), the difference between congruous and incongruous final words was delayed as compared with incongruities with mismatching initial phonemes. Both results can be accomodated by the view that those aspects of lexical processing reflected by the N400 are triggered at the point when a word can uniquely be identified. But both results are also consistent with the view that the onset latency of the word–pseudoword and sentence congruity effects reflects the time at which the stimulus became semantically incoherent.

These alternatives generate contrasting predictions evaluated here by four main experimental conditions that differed in the relationship between the meaning of a sentence frame and the phonemic structure of the sentence-terminal word. Table 2 shows two types of sentence-final words used as anchor points: *Cohort congruous* words were meaningful sentence completions, and *plain incongruous* words were semantically anomalous sentence completions with no special phonological relationship to the congruous completions of those sentences. In two other conditions, the final words were semantically anomalous but phonemically similar to the expected sentence completion. *Cohort incongruous* words shared initial phonemes with the congruous sentence completions (e.g., DOLLARS when DOLPHINS was expected). *Rhyme* words shared final phonemes (e.g., MUFFINS when DOLPHINS was expected). If semantic integration begins after a word can uniquely be identified from acoustic input alone, the ERP congruity effect will begin shortly after the IP. Because the cohort incongruous and rhyme words have similar IPs (30 ms apart, see Table 1), both of these conditions should diverge from congruous completions at about the same time. Different predictions emerge if seman-

Table 2
Sample Stimuli in Experiment 2

Cohort congruous/cohort incongruous/rhyme
It was a pleasant surprise to find that the car repair bill was only seventeen *dollars/dolphins/scholars.*
Most marine mammals have some sort of legal protection, but fishermen continue to kill *dolphins/dollars/muffins.*
The referee got tired of the coach's behavior and gave his team a *penalty/pendulum/loyalty.*
The visitor stared at the grandfather clock and seemed almost hypnotized by the swinging of the *pendulum/penalty/asylum.*
The movie was meant to be a horror flick, but the acting was so bad it was *funny/funnel/penny.*
He spilled some of the oil onto the engine because he didn't have a *funnel/funny/tunnel.*
Instead of sending flowers, they asked everyone to make a donation to their favorite *charity/chariot/cavity.*
The Roman general made his appearance in battle gear, with a white horse pulling his golden *chariot/charity/idiot.*
Sir Lancelot spared the man's life when he begged for *mercy/mermaid/fancy.*
The old sailor kept a straight face as he told of his love affair with a beautiful *mermaid/mercy/grenade.*
Plain congruous
The mill worker caught his hand in a piece of machinery and was rushed to the *hospital.*
Plain incongruous
The gold medal winner from Brazil started out slow but took the lead in the second *bureau.*

tic integration begins with only partial information about word identity. If semantic processing follows the acoustic input continuously, the congruity effect may begin before the IP, when word-initial information does not match semantic expectations. Under this scenario, the rhyme condition will show an earlier congruity effect than will the cohort incongruous condition. The rhyme condition also provides a control for general phonemic similarity between congruous and incongruous completions.

Method

Participants. Fourteen women and 7 men served as paid volunteers. All were native English speakers who reported normal hearing, normal vision, and no history of neurological disorder. Their mean age was 26.6 years (range, 19–39). Twenty were right-handed, 1 was left-handed, and 2 of the right-handers reported a left-handed parent or sibling. Nine were college students, and 12 had some postgraduate education: All were naive as to the purpose of the experiment. The data from 2 additional participants were not analyzed: 1 failed to return for the second session, and 1 other's data were lost in a computer malfunction.

Materials. The stimulus materials consisted of spoken sentence frames paired with final words that formed semantically congruous or incongruous endings. Sample stimuli are shown in Table 2. The congruous conditions included sentences ending with the cohort words described in Experiment 1 (*cohort congruous*), and additional sentences added to balance the number of congruous and incongruous sentences presented during the experiment (*plain congruous*). The incongruous final words fell into three conditions: phonemically dissimilar to the congruous ending for that sentence (*plain incongruous*), similar in initial phonemes (*cohort incongruous*), or similar in final phonemes (*rhyme*). The sentence-final words in the cohort and rhyme conditions were those used in Experiment 1, and the audio files were the full-duration versions of those previously subjected to gating (see Table 1 for descriptive statistics of the cohort and rhyme words).

The plain congruous and plain incongruous final words were "plain" merely because they were not included in the previous gating experiment and were not paired with phonemically similar words in Experiment 2 (although like all words, they do have phonemic neighbors). The plain incongruous condition served as an anchor point to evaluate N400 amplitude when a sentence completion was both semantically anomalous and phonemically dissimilar to the unpresented congruous completion. The plain congruous sentences were true fillers, included to make the proportion of congruous and incongruous sentences in the experiment more nearly equal despite the fact that the majority of the experimentally relevant sentences were incongruent. The plain final words were recorded, digitized, and edited using the procedures described in Experiment 1. The plain final words were roughly equated in frequency and duration to the cohort and rhyme words (frequency, $M = 85$ ms; duration, $M = 585$ ms).

The sentence frames were spoken and recorded without their final words and were then digitized and edited to yield one audio file for each sentence frame. Excluding the final words, sentences averaged 12.3 words (range, 6–29) and 3.7 s in duration (range, 1.665–7.887). A separate group of 160 college students participated in a cloze procedure to establish the predictability of the sentence-final words in the cohort congruous condition; each individual participant was given 70 or 71 sentences with the final word replaced by a blank. Each sentence was completed by a minimum of 30 participants for course credit. This procedure indicated that the cohort congruous completions had a mean cloze probability of 58%, with a range of 0% to 100%. The plain congruous sentences used as fillers in this experiment were drawn from previously normed sets of stimuli; their final words had a cloze probability of 61% (range, 4%–100%).

The set of critical stimuli consisted of 176 "units" consisting of two sentence frames and three possible completions for each frame (Table 2). From these, three stimulus lists were constructed such that no participant heard a sentence frame or a final word more than once. Across subjects, each sentence frame was paired with cohort congruous, cohort incongruous, and rhyme completions in equal proportions. Each cohort word was presented equally often as a congruous and an incongruous sentence completion. Each participant listened to 117[4] sentences each in the cohort congruous, cohort incongruous, and rhyme conditions, with equal numbers of high- and low- (relative) frequency final words in the two cohort conditions. All stimulus lists additionally included the same set of 136 plain congruous and 44 plain incongruous sentences, for a total of 531 sentences split between two recording sessions. The five conditions were randomly intermixed within each stimulus list.

Behavioral task. We used a behavioral task that mandates close attention to the stimulus materials but that does not require any decisions while listening to the sentences. After each sentence, a target word was presented visually, and participants indicated whether or not the word had occurred in the preceding sentence. For each sentence condition, half of the targets were part of their associated sentences, and half were not. Final words were never used as targets. Because the exact target for a given sentence was not predictable in advance, this task avoids any confound between the experimental manipulations of interest and the speed or accuracy of task performance. Previous research has shown that the sensitivity of the N400 component to semantic context is relatively independent of assigned task but may be obscured by a temporally overlapping component (the P300), which is sensitive to decision-related factors across a variety of linguistic and nonlinguistic paradigms (see Kutas & Hillyard, 1989; Kutas & Van Petten, 1994). Accuracy in the target task thus serves as a general check on participants' alertness, but it is not otherwise of interest in the present study.

Procedure. Each individual participated in two sessions lasting about 3 hr each. The Peabody Picture Vocabulary Test—Revised (PPVT–R; Dunn & Dunn, 1981) was administered at the beginning of the first session, and the Reading Span Test (Daneman & Carpenter, 1980) was administered at the start of the second session.

Participants were comfortably seated facing a table with audio and video monitors under the control of a PC computer. Volume level for the audio monitor was adjusted to each person's comfort at the start of the experiment. The video monitor continuously displayed a rectangular box; participants were instructed to maintain eye fixation on the box during each sentence. A single trial consisted of a sentence frame, 500 ms of silence, and a final word, followed by a visual target presented 3 s after the onset of the final word. Keypresses with the right and left index fingers were used for yes and no responses to the target words, with the mapping between response and right or left hand counterbalanced between the two sessions. The intertrial interval was 7 s, with brief rest periods every 30 trials.

Electrophysiological methods. The electroencephalogram (EEG) was recorded with tin electrodes mounted in a commercially available elastic cap. Midline central (Cz) and parietal (Pz) recording sites were used, along with lateral pairs of electrodes over the posterior temporal (T5 and T6) and occipital (O1 and O2) scalp as defined by the 10–20 system (Jasper, 1958). Three additional lateral pairs were used: (a) a frontal pair placed midway between F7–8 and T3–4 (approximately over Broca's area and its right hemisphere homolog, Bl and Br, respectively), (b) a temporoparietal pair placed 30% of the interaural distance lateral and 12.5% of the inion–nasion distance posterior to Cz (approximately over Wernicke's area and its right-hemisphere homolog, Wl and Wr, respectively), and (c) an additional temporal pair placed 33% lateral to Cz (Tl and Tr). Each scalp site was referred to an off-line average of the left and right mastoids (see Van Petten & Kutas, 1988). Vertical eye movements and blinks were monitored via an electrode placed below the right eye referred to the left mastoid. Horizontal eye movements were monitored via a right-to-left bipolar montage at the external canthi.

The EEG was amplified by a Grass Model 12 polygraph (Grass Instruments, West Warwick, RI) with half-amplitude cutoffs of 0.01 and 100 Hz, digitized on-line at a sampling rate of 170 Hz and stored on optical disk along with stimulus codes for subsequent averaging. Trials with eye movement, muscle, or amplifier blocking artifacts were rejected before averaging.

Responses to the final words were averaged with two timelock points: the onset of the word (all conditions) and the IP established in Experiment 1 for the cohort and rhyme conditions. The averaging epoch included a 1,500 ms baseline preceding the timelock point and 1,500 ms of poststimulus activity.

Results and Discussion

Behavioral tests. The mean score on the PPVT–R was 117 (SE = 3.0); mean reading span was 3.0 (SE = 0.2). Splitting the participant group into those with higher or lower scores on these measures yielded no observable differences in the results presented later. Accuracy in the postsentence probe recognition task averaged 91% for words that occurred in the sentence (SE = 0.7) and 94% for words that did not (SE = 0.3). The high accuracies indicate that participants were indeed attentive to the sentences; there was too little variability to examine relationships between accuracy and the ERP measures.

General characteristics of the event-related potentials. Figure 4 shows the ERPs elicited in the plain sentence conditions, for which the averaging process was timelocked to the onset of the final words. Before the onset of the final word, the epoch includes the last 1,000 ms of the sentence frame (−1,500 to −500 ms). Although words were presented in this time period, this portion of the epoch includes no clear ERPs. The onsets of these sentence-intermediate words occurred at variable times with respect to the timelock point; such latency variability leads to a smearing of positive-going and negative-going components in the averaged waveform. Early components of the auditory ERP (those elicited by all stimuli, regardless of cognitive significance) are also known to have long refractory periods such

---

[4] The total number of sentences in these conditions did not divide evenly into three lists and varied between 117 and 118. The number of high- and low-frequency final words also varied by one item across lists. Note that we did not use a simple Latin square design but rather maximized the number of trials in each condition while maintaining the constraints noted in the text. A stimulus unit consisted of two sentence frames in which the congruous completions shared initial phonemes—Frames A and B (see examples in Table 2). In Session 1, each participant heard half of the A frames and half of the B frames, with the other half delivered in Session 2. Within a session, both the A and B frames were completed equally often with a cohort congruous, cohort incongruous, or rhyme word. If a congruous completion of the A frame was heard in Session 1, then a rhyme completion of the B frame was heard in Session 2, and vice versa if a rhyme completion was heard in Session 1. If a cohort incongruous completion of the A frame was heard in Session 1, then a cohort incongruous completion to the B frame was heard in Session 2. In addition to the constraints noted in the text, this counterbalancing scheme ensured that neither cohort pairs of words (DOLLAR–DOLPHIN) nor rhyming pairs (DOLLAR–SCHOLAR) were presented together in an experimental session.

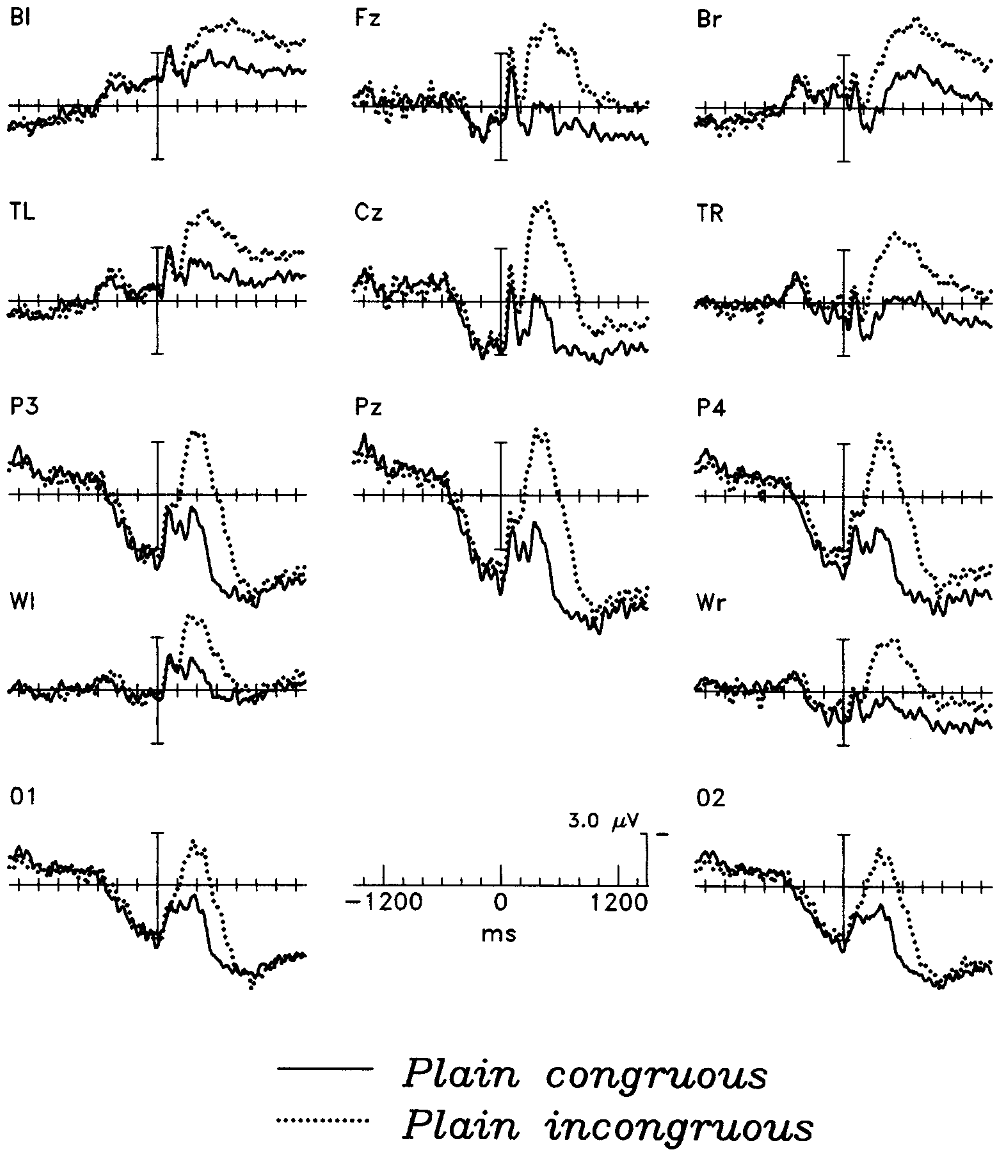

Figure 4. Grand average event-related potentials from 21 participants, elicited by the plain congruous and plain incongruous sentence-final words. Note that negative is plotted up in this and subsequent figures. Electrode sites on the left side of the head are shown in the left column, midline sites in the middle column, and right scalp sites in the right column. Top to bottom of the figure corresponds with anterior to posterior across the scalp. Time 0 is the onset of sentence-final words. "F" denotes frontal, "C" central, "P" parietal, and "O" occipital scalp sites. "B" ("Broca's") denotes scalp sites over inferior prefrontal brain regions, and "W" ("Wernicke's") scalp sites roughly over the temporal–parietal junction. "L" and odd numbers indicate the left side of the head; "R" and even numbers indicate the right side of the head.

that they are much reduced in amplitude during continuous speech with few silent intervals (Davis, Mast, Yoshie, & Zerlin, 1966).

The offset of each sentence frame was fixed at 500 ms before final-word onset. Figure 4 shows a clear positive wave during this brief silent pause (−500 ms–0 ms), particularly at parietal scalp sites. The parietal scalp distribution of the "pause positivity" is characteristic of the P300 family of responses to unexpected events. Previous studies have demonstrated that P300s are emitted when stimuli are omitted from a train of clicks occurring at regular intervals (Sutton, Tueting, Zubin, & John, 1967) or when pauses are unpredictably inserted into spoken sentences or musical melodies (Besson & Faita, 1995; Besson, Faita, Czternasty, & Kutas, 1997). Within-sentence pauses of a half second or longer are common in spontaneous speech, but rarely after the penultimate word (Dechert & Raupach, 1980; Deese, 1984; Goldman-Eisler, 1972). The observation of a P300 in response to the break between the sentence frames and final words indicates that listeners were sensitive to the rupture of sentence prosody induced by the pause, even though every sentence contained one.

At frontal, central, and anterior temporal sites, the sentence-final words elicited an N100 potential typical of the auditory ERP response to stimuli with an abrupt acoustic onset. Our primary interest is in the subsequent N400 component and its sensitivity to sentence congruity and the phonological relationships between congruous and incongruous words.

Congruous versus incongruous sentence completions. Figure 4 shows that the plain incongruous sentence completions elicited a much larger N400 component than did the plain congruous completions, beginning at 200 ms or less after word onset, and extending to at least 1,000-ms postonset. The contrast between these two conditions replicates other reports (Connolly et al., 1990; Holcomb & Neville, 1991; McCallum et al., 1984). Although our primary interest is in the latency of the N400 congruity effect across conditions, we initially quantified the expected result of larger N400s to incongruous endings. This initial measurement took mean amplitude in a broad latency window from 200- to 800-ms poststimulus onset using the 1,500 ms of activity preceding final word onset as a baseline. For the plain sentence types, an ANOVA with factors of congruity and scalp site (13 levels) indicated that the basic congruity effect was a robust one, $F(1, 20) = 95.00$, $MSE = 14.60$, $p < .0001$. Similarly, the cohort incongruous and rhyme sentence completions elicited larger N400s than did the cohort congruous completions: cohort incongruous, $F(1, 20) = 57.10$, $MSE = 7.01$, $p < .0001$; rhyme, $F(1, 20) = 116.40$, $MSE = 8.64$, $p < .0001$.

Relevance of final phonemes: The rhyme condition. Figure 5 indicates that the rhyme condition elicited ERPs very similar to the plain incongruous condition, which incorporated no special phonemic similarity to congruous completions. Phonemic similarity between the expected congruous completion of a sentence and the incongruous word actually presented did not influence N400 amplitude, when the shared phonemes occurred at the ends of words. If the rhyming manipulation were effective, we would expect it to be apparent relatively late in the recording epoch, during (or after) the presentation of the final syllable. We thus compared the rhyme and plain incongruous conditions for successive 100-ms portions of the waveform, from 200- to 1,000-ms poststimulus onset. The null results of these contrasts are shown in Table 3. Given this outcome, we consider the plain incongruous and rhyming conditions as functionally equivalent in this experiment. Because IPs for the rhyming words were defined in the gating study (which did not include the plain incongruous words), in some of the subsequent analyses we used the rhyme condition as a comparison for the cohort conditions.

It was a bit surprising to find no effect of the rhyme manipulation, given that informal debriefing of the participants indicated that they had indeed noticed that some of the terminal words rhymed with the expected completion. The absence of any difference between the rhyme and the plain incongruous items suggests that when participants were aware of the rhymes, the realization occurred beyond the recording epoch (1.5 s postword onset). A handful of behavioral studies have used rhyme manipulations to address the question of whether spoken word identification proceeds strictly from left to right such that word-initial phonemes rule out any mismatching lexical candidates. Strict left-to-right processing was a tenet of early versions of Marslen-Wilson's cohort model (Marslen-Wilson & Welsh, 1978), which was criticized for its inability to allow recovery from mispronunciations or noise at word onsets (e.g., the ability to recognize SHIGARETTE as a token of CIGARETTE; Grosjean, 1985; Marslen-Wilson & Welsh, 1978). This tenet has been relaxed in more recent versions of the model, to allow a limited role for "late entry" of words with mismatching initial phonemes into the set of possible candidates (Marslen-Wilson, 1993). Some of the empirical results that forced this revision have come from experiments starting with semantically related word pairs, but then altering the first phoneme of the "prime" word to yield pairs such as FEATURE–STUDENT. Such word pairs yield small facilitations in lexical-decision time, but only if the prime word (or nonword) is very similar to the original related item, differing in a single phoneme or even one phonetic feature (Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson & Zwitserlood, 1989). The rhyme manipulation of the present study is formally similar to these experiments in that semantically congruent words were replaced with items that shared final phonemes. However, our stimuli differed considerably from the lexical-decision experiments in that the shared final phonemes were often preceded by more than one mismatching initial phoneme (e.g., PENALTY–LOYALTY). Although the absence of a rhyme effect in the current results indicates that listeners were not immediately reminded of PENALTY when they heard LOYALTY, this finding bears only minimally on the issue of how listeners recover from mispronounced initial phonemes. The null rhyming effect places stronger constraints on the question of how listeners processed the experimental sentences, a point taken up in the General Discussion section.

Relevance of initial phonemes: The cohort conditions. Semantically incongruous words sharing initial phonemes with a congruous sentence completion elicited a larger N400 than did congruous completions, but only after a substantial delay. Figure 6 shows that the sentence congruity effect was delayed by some 200 ms in the cohort incongruous condition as compared with the rhyme condition. When word-initial phonemes were appropriate for the sentence context (e.g., DOLLARS when DOLPHINS would have been a congruous completion), the congruity effect began at about 350-ms poststimulus onset. In contrast, when word-initial phonemes could not form a congruous completion (e.g., MUFFINS instead of DOLPHINS), the incongruous completions diverged from the congruous at about 200 ms. The latency shift was initially quantified by measuring peak N400 latency—the timing of the most negative point in the latency window of 200–800 ms. Peak latencies were 525 ms for the

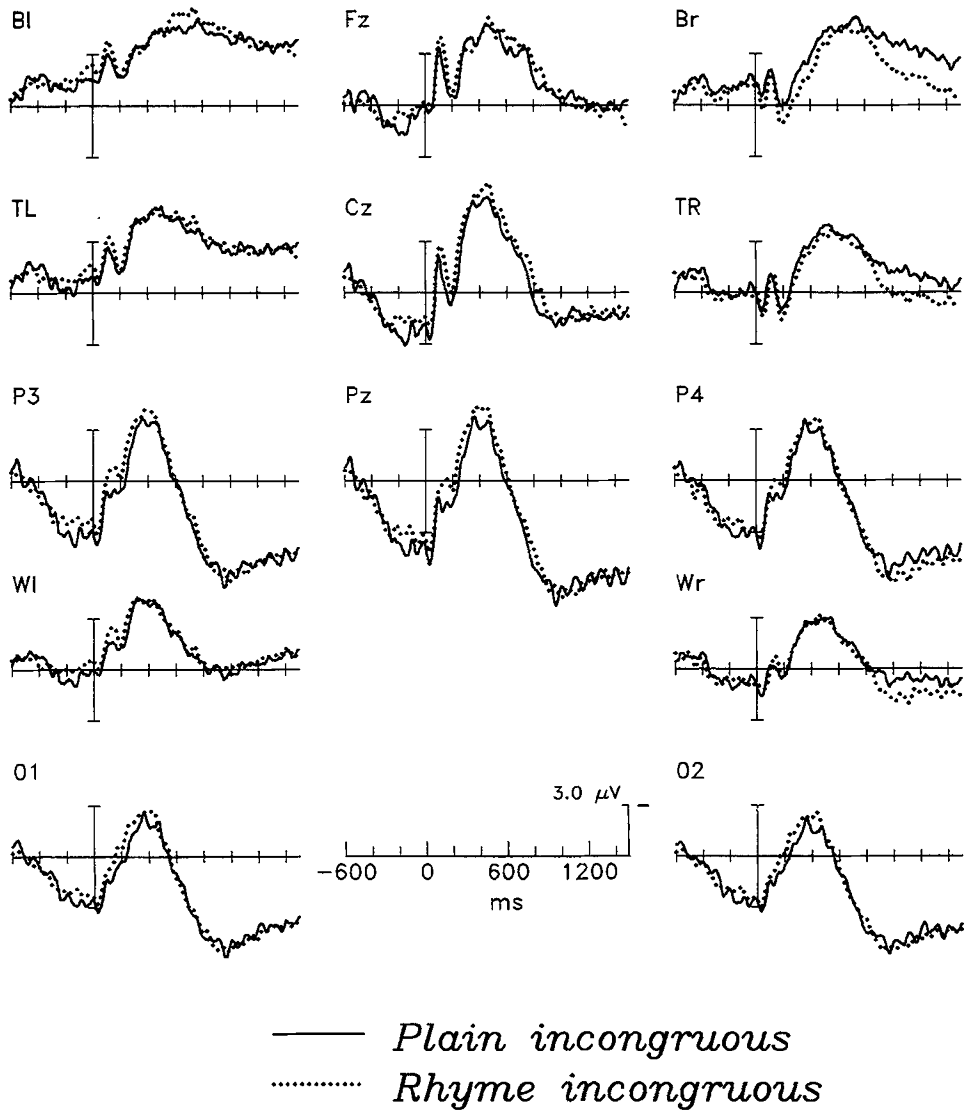

Figure 5. Grand average event-related potentials elicited by the plain incongruous and rhyme sentence-final words. Time 0 is the onset of sentence-final words. "F" denotes frontal, "C" central, "P" parietal, and "O" occipital scalp sites. "B" ("Broca's") denotes scalp sites over inferior prefrontal brain regions, and "W" ("Wernicke's") scalp sites roughly over the temporal–parietal junction. "L" and odd numbers indicate the left side of the head; "R" and even numbers indicate the right side of the head.

cohort incongruous condition, 469 ms for the rhyme condition, and 471 ms for plain incongruous condition: cohort incongruous versus rhyme, $F(1, 20) = 15.60$, $MSE = 27{,}321$, $p < .001$; cohort incongruous versus plain incongruous, $F(1, 20) = 16.40$, $MSE = 23{,}914$, $p < .001$. This result indicates that the latency shift was statistically reliable, but there is little reason to assume that the moment an ERP component reaches peak amplitude has any special significance. Of greater interest is when the incongruous conditions first began to differ from the congruous condition. A more detailed temporal analysis was performed by comparing conditions in successive 100-ms latency windows, as shown in Table 4. The congruity effects were not significant during the first 200 ms after word onset. The rhyme condition first began to diverge from the cohort congruous condition in the 200- to 300-ms latency window. In contrast, the cohort incongruous condition did not begin to differ from the cohort congruous sentences until the 400- to 500-ms window.

Semantic integration and the isolation point. The preceding analyses were based on averages time locked to word onset, across sets of words with varying IPs. Those results

Table 3
Rhyme Versus Plain Incongruous

Latency window (ms)	$F(1, 20)$	*MSE*	$p<$
200–800'	0.89	6.80	*ns*
200–300	3.48	8.31	*ns*
300–400	0.01	10.7	*ns*
400–500	0.49	10.6	*ns*
500–600	0.36	11.5	*ns*
600–700	0.16	14.3	*ns*
700–800	0.72	13.5	*ns*
800–900	0.16	13.9	*ns*
900–1,000	0.58	14.3	*ns*

Note. Results of analyses of variance on mean amplitude. Latency window is in milliseconds after onset of sentence-final word, with respect to 1,500 ms of prestimulus activity as a baseline. Scalp site was used as a factor (13 levels) but did not qualify the main result of no significant difference between plain incongruous and rhyme words.

suggest that the IP is not the critical time point for elicitation of the N400. Although the cohort and rhyming words had only slightly different average IPs (376 vs. 345 ms), cohort and rhyme incongruities differed by 200 ms in the onset of the N400 sentence congruity effect. However, a better procedure for examining the relationship between the time course of semantic integration and the IP is to correct for the variability in IP across individual words. Figure 7 shows ERPs in which the IP was used as the timelock signal for the averaging process. Note that the N100 and other sensory components of the auditory ERP are absent in these waveforms because these components are triggered by acoustic onsets, but the ERPs are timelocked to a point in the middle of each word.

Onset latencies of the congruity effects were evaluated by measuring successive 100-ms latency windows, beginning 500 ms before the IP (relative to a baseline of -1500 to -500 ms). No significant congruity effects were observed earlier than 200 ms before the IP, as shown in Table 5. For words with shared initial phonemes, the first divergence between congruous and incongruous completions occurred very close to the IP, in the 0- to 100-ms window. When initial phonemes did not match the expected word, the congruity effect preceded the IP by some 200 ms.

Figure 7 also shows that although the congruity effect for cohort words began later than for rhyme words, it did not persist for longer (see also Table 5). In other words, the distinction between words with appropriate and inappropriate initial phonemes did not take the form of an overall latency shift for the onset, peak, and offset of the congruity effect. Instead, the initial 200 ms of the congruity effect evident in the rhyme condition was simply absent in the cohort incongruous condition. In the General Discussion section, we interpret this pattern of latencies as reflective of a continuous mapping from acoustic input to semantic representations.

An alternative interpretation of the different durations of the congruity effects in the cohort incongruous versus rhyme conditions lies in considering the "early" (pre-IP) and "late" portions of the congruity effect as indexes of qualitatively distinct processes or as two separate ERP components. Connolly and colleagues have argued for the existence of a *phonological mismatch negativity* (PMN), which reflects the discrepancy between expected and presented initial phonemes and separable from the N400 per se (Connolly & Phillips, 1994; Connolly et al., 1992). Because the PMN is reported to have the same sensitivity to sentence context as the N400, these investigators base the claim of distinct components primarily on the same sort of result as that observed here: delayed onset of the congruity effect when initial phonemes match the congruous ending. The PMN has also been described as having the same spatial distribution across the scalp as the N400, which makes it particularly difficult to distinguish between a single component with variable latency and two components.

Connolly and colleagues (Connolly & Phillips, 1994; Connolly et al., 1992) have suggested that the separation between PMN and N400 is best observed in the ERPs of single participants, because even a small degree of latency variability between subjects can smear the two components into what appears to be a single wave when viewed in the grand average across subjects. We thus examined single-participant averages for signs of a double peak in the congruity effect. The prediction from a two-component model is that the rhyme condition should include two peaks reflecting initial phonological mismatch and subsequent purely semantic processing, whereas the cohort incongruous condition should include only one. Figure 8 displays these two congruity effects in the first 12 participants, who are representative of the full sample. Each panel contains a difference waveform created by a point by point subtraction of the congruous completions from the cohort incongruous and rhyme conditions. It can be seen that many of the individuals do display subpeaks within the overall congruity effects. However, it is also clear that there is a pronounced 10-Hz rhythm throughout the waveforms and that the rhyme difference wave is no more likely to contain multiple peaks than the cohort incongruous difference wave. The 10-Hz activity is undoubtably a residue of the alpha rhythm in the spontaneous EEG, which has not been fully eliminated by averaging multiple trials of EEG into an ERP. In any case, neither the grand average (Figures 4–7) nor single-participant ERPs display any convincing evidence for multiple components of the sentence congruity effect.

Influences of cloze probability and contextual constraint. Across the full stimulus set, the cloze probability of the congruous final words was 58%. The congruous words were thus not fully predictable, yet the sentence frames provided moderately strong support for these items. The range of cloze probabilities across the stimulus set was broad enough to allow evaluation of this factor. Congruous words were split into high- versus low-cloze probability categories using 60% as the threshold. This procedure yielded 166 high-cloze congruous completions (mean probability = 86%) and 191 low-cloze (mean probability = 36%) in the full stimulus set. Each participant heard roughly one third of these high- and low-cloze subsets, given the counterbalancing constraints described in the *Method* section.

Both the cohort incongruous and rhyme endings were semantically anomalous and had cloze probabilities of zero.

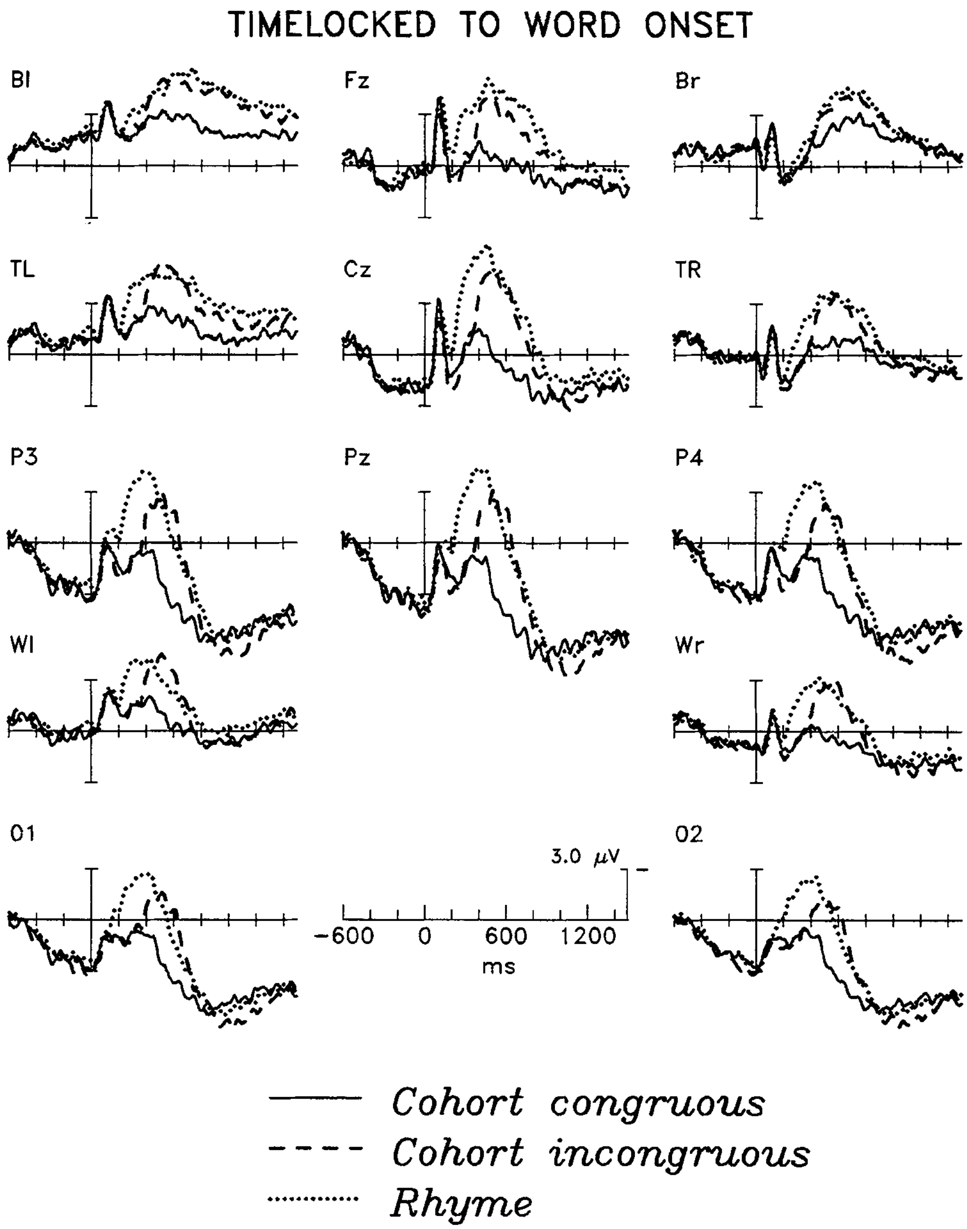

Figure 6. Grand average event-related potentials elicited by the cohort congruous, cohort incongruous, and rhyme sentence-final words, timelocked to word onset. Time 0 is the onset of sentence-final words. "F" denotes frontal, "C" central, "P" parietal, and "O" occipital scalp sites. "B" ("Broca's") denotes scalp sites over inferior prefrontal brain regions, and "W" ("Wernicke's") scalp sites roughly over the temporal–parietal junction. "L" and odd numbers indicate the left side of the head; "R" and even numbers indicate the right side of the head.

However, it is of interest to determine if the N400 elicited by these incongruous words is influenced by the strength of the preceding sentence frame. The *contextual constraint* of a sentence frame can be defined as the cloze probability of its most preferred completion, so that a frame such as "The photo was printed in black and ___" is of high constraint because most readers select *white* as a completion, whereas a frame such as "She was soothed by the gentle ___" is of low constraint because there are many acceptable endings with none being strongly preferred. In the visual modality, the cloze probability of the word actually presented has a powerful influence on N400 amplitude, but the contextual constraint of the sentence frame does not (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). In other words, anomalous words are equally anomalous whether they occur after strong or weak semantic constraints. Similarly, weakly

Table 4
Time Course of the Sentence Congruity Effects, Measured From Word Onset

Latency window (ms)	Cohort congruous vs. rhyme			Cohort congruous vs. cohort incongruous		
	$F(1, 20)$	*MSE*	$p<$	$F(1, 20)$	*MSE*	$p<$
0–100	0.10	10.7	*ns*	0.22	6.0	*ns*
100–200	3.31	7.2	*ns*	1.31	9.7	*ns*
200–300	47.4	12.5	.0001	0.74	12.4	*ns*
300–400	96.4	11.5	.0001	1.43	7.7	*ns*
400–500	146.1	10.0	.0001	55.4	13.7	.0001
500–600	137.8	10.9	.0001	131.9	13.4	.0001
600–700	67.3	14.3	.0001	65.6	18.2	.0001
700–800	40.7	15.3	.0001	16.5	15.3	.001
800–900	13.6	20.2	.002	3.52	20.1	(.10)
900–1,000	5.6	12.3	.05	0.34	20.2	*ns*

Note. Results of analyses of variance in which sentence type and electrode site were used as repeated measures: main effect of congruity. Latency window is in milliseconds after onset of sentence-final word.

predictable congruent words elicit N400s of equivalent (moderate) amplitude whether they occur after a sentence frame that strongly predicts some other word or a sentence frame with no single strong completion. However, experiments in the visual modality rarely yield any variability in the latency of the N400 (cf. Van Petten & Kutas, 1987), so that it will be of some interest to determine whether the responses to anomalous words begin earlier when the sentence frames encourage a strong expectation for some other word. The sentence frames were accordingly split into high versus low constraint, depending on the probability of the most common completion supplied by the cloze participants (threshold of 60%). This procedure yielded 176 sentences in each category, with mean constraints of 85% and 38% for the high and low categories, respectively.

Figure 9 displays ERPs timelocked to the IP. The noise level of these data is higher than those of previous figures, given that they are based on half the number of trials. Nonetheless, a small effect of cloze probability is evident in the responses to congruent completions. The low-cloze final words elicited somewhat larger N400s than high-cloze endings, as previously reported in both the visual and auditory modalities (Connolly, Phillips, & Forbes, 1995; Connolly et al., 1992; Kutas & Hillyard, 1984). Analyses of successive 100-ms latency windows revealed a significant cloze effect in a brief time span just before and after the IP: −100 ms–0 ms, $F(1, 20) = 4.79$, $p < .05$, $MSE = 25.39$; 0–100 ms, $F(1, 20) = 6.14$, $p < .05$, $MSE = 28.21$. In contrast to this small but reliable influence of cloze probability, contextual constraint had no influence on the responses to anomalous words. Figure 9 shows that the sentence congruity effect for words with inappropriate onsets—the rhyme condition—began well before the IP, regardless of how strongly the sentence frame predicted a single alternate completion.

Experiment 3

The design of Experiment 2 included one feature that was somewhat obtrusive to natural listening conditions. We included a 500-ms silent period between the termination of a sentence frame and onset of its final word. Although midsentence pauses of this duration are not uncommon in natural speech, they rarely occur in this sentence position, nor do most speakers pause at the same point in every sentence they utter. The half-second silence was designed to allow brain activity elicited by the sentence frame to die off and thus yield more isolated observation of brain activity elicited by the critical final word without the complication of overlapping responses to the sentence frame and final word. However, the results of Experiment 2 suggest that this precaution may have been unnecessary. Examination of Figure 4 indicates little observable ERP activity during the sentence frame (−1,500 to −500 ms in that figure). The individual words are so poorly timelocked to the termination of the sentence frame that the sequence of positive and negative potentials elicited by each word yields fairly good cancellation in the average across multiple sentences. Given this result, it appears technically feasible to extract ERPs elicited by spoken words that are more immediately preceded by running speech.

Experiment 3 was designed to evaluate the generality of the previous results by reducing the duration of the silent pause preceding final-word onset. Elimination of the half-second pause serves to obviate the possibility that listeners engage in special attempts to predict sentence-final words during this time. Comprehension of natural speech may well involve some degree of both overt prediction, as well as less specific expectations about upcoming words, but laboratory timing parameters that more closely approach those of natural speech are more clearly generalizable to everyday listening.

Method

Participants. Ten men and 8 women served as paid volunteers. All were native English speakers who reported normal hearing, normal vision, and no history of neurological disorder. Their mean

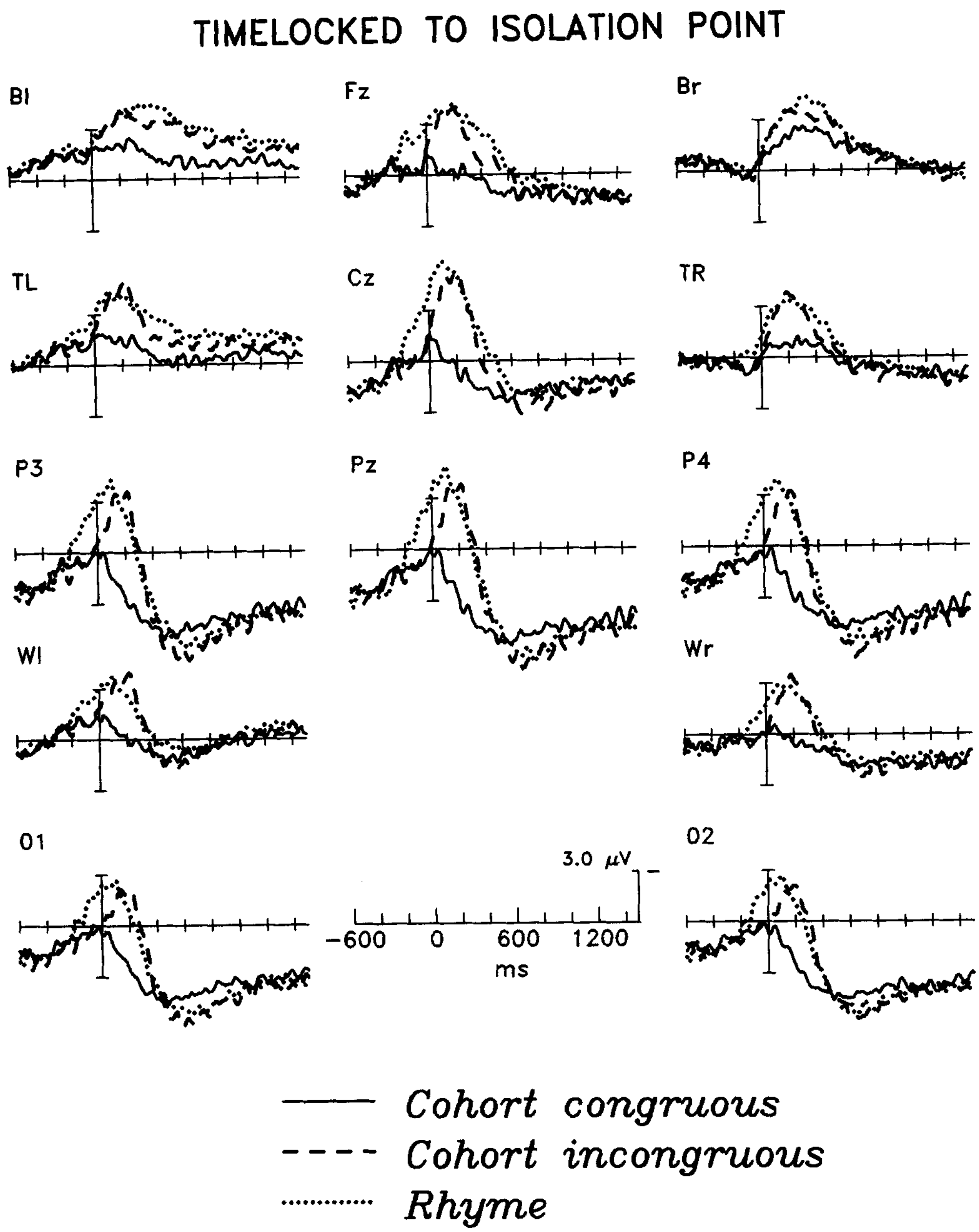

Figure 7. Grand average event-related potentials elicited by the cohort congruous, cohort incongruous, and rhyme sentence-final words, timelocked to the isolation point Time 0 is the isolation point for the sentence-final words, established in Experiment 1. "F" denotes frontal, "C" central, "P" parietal, and "O" occipital scalp sites. "B" ("Broca's") denotes scalp sites over inferior prefrontal brain regions, and "W" ("Wernicke's") scalp sites roughly over the temporal–parietal junction. "L" and odd numbers indicate the left side of the head; "R" and even numbers indicate the right side of the head.

age was 26.2 years (range, 21–34). Fourteen were right-handed, 5 were left-handed, and 6 of the right-handers reported a left-handed parent or sibling. Ten were college students, 8 had completed a bachelor's degree, and 5 of these had some postgraduate education. Mean reading span was 3.7 (SE = 0.2); mean score on the Peabody Vocabulary Test (Dunn & Dunn, 1981) was 116 (SE = 3).

Stimuli and procedure. The stimulus materials and experimental procedures were identical to those of Experiment 2, with the exception that sentence-final words followed the offset of sentence frames after 50 ms of silence rather than 500 ms.

Results

The left column of Figure 10 displays the ERPs elicited by the cohort congruous, cohort incongruous, and rhyme condi-

Table 5
Time Course of the Sentence Congruity Effects, Measured From Isolation Point

Latency window (ms)	Cohort congruous vs. rhyme			Cohort congruous vs. cohort incongruous		
	$F(1, 20)$	*MSE*	$p<$	$F(1, 20)$	*MSE*	$p<$
−500 to −400	0.12	5.9	*ns*	0.85	6.5	*ns*
−400 to −300	0.46	8.0	*ns*	0.13	4.2	*ns*
−300 to −200	0.14	7.6	*ns*	0.29	9.9	*ns*
−200 to −100	11.1	15.1	.005	0.56	10.0	*ns*
−100 to 0	36.3	11.9	.0001	0.22	10.5	*ns*
0 to 100	54.3	20.4	.0001	18.5	12.6	.0005
100 to 200	96.3	18.3	.0001	111.8	14.5	.0001
200 to 300	99.0	14.3	.0001	70.5	25.0	.0001
300 to 400	51.6	14.1	.0001	17.2	24.1	.0005
400 to 500	16.7	17.5	.001	2.43	14.7	*ns*
500 to 600	8.23	12.6	.01	0.25	16.3	*ns*
600 to 700	0.82	9.8	*ns*	1.92	15.7	*ns*
700 to 800	0.17	12.1	*ns*	1.51	18.0	*ns*

Note. Results of analyses of variance in which sentence type and electrode site were used as repeated measures: main effect of congruity. Latency window is in milliseconds after isolation point for sentence-final word.

tions, timelocked to the acoustic onset of these sentence-final words. Independent of the experimental conditions, comparison with Figure 6 (the analogous conditions from Experiment 2) reveals two global differences due to the between-experiment timing manipulation. First, the amplitude of the auditory N1 is very small in Experiment 3. The sensitivity of the N1 to interstimulus interval is well documented; its small amplitude here is much like prior studies using tones or clicks with brief temporal separations (Davis et al., 1966). Second, the epoch preceding final-word onset includes no observable ERPs, in contrast to the positive component observed in the previous experiment. In Experiment 2, we dubbed this parietal positivity the *pause positivity* and assumed that it was elicited by the silent period separating the penultimate from the sentence-final word or the prosodic break created by splicing sentence frames with final words that were not originally spoken with sentence-ending intonation. The absence of a pause positivity before word onset in Experiment 3 indicates that this speculation was correct, but it cannot indicate which of the two factors is more critical. If temporal rupture were the critical eliciting event, we would expect the pause positivity to be reduced or absent in Experiment 3, given the brevity of the silent period. If the positivity instead reflected the prosodic break, it would be timelocked to the break point (at −50 ms in Figure 10's left column) and thus occur within the same time frame as activity elicited by the sentence-final word. The difficulty of resolving ERPs elicited by temporally overlapping events makes it impossible to determine whether the pause positivity is present or absent in Experiment 3.

Amplitude analyses of the central experiment manipulations in Experiment 3 were conducted in the same way as the previous experiment and yielded similar results. Within a broad latency window of 200- to 800-ms poststimulus onset, plain incongruous, cohort incongruous, and rhyme completions all elicited larger N400s than the cohort congruous completions: $Fs(1, 17) = 43.20$, 20.80, and 35.00 respectively, all $ps < .0005$. The rhyme condition did not differ from the plain incongruous condition (*F*s for successive 100-ms latency windows < 2.69).

Figure 10 indicates that the latency results also replicated those of Experiment 2. Relative to word onset, the N400 sentence congruity effect for rhyme completions began at about 200 ms, whereas the N400 elicited by cohort incongruous completions was delayed until about 350 ms—about the same as the IP for these words (see Table 6 for statistical analyses comparable to those of Table 4). Relative to the IP, the congruity effect in the cohort incongruous condition began at about Time 0, whereas the effect for rhyme completions preceded the IP by some 200 ms (see Table 7).

General Discussion

The gating results from Experiment 1 demonstrated both continuous and discontinuous processes in spoken word identification. Listeners generated progressively fewer candidates as they heard longer fragments of the target words, and this reduction in cohort size was linear—up to a point. At the critical duration defining the IP, accuracy in identifying the target word increased abruptly, and the number of alternative candidates dropped off steeply. The central finding of Experiments 2 and 3 was that the time course of semantic integration was not invariably linked to that of successful word identification. Instead, semantic context influenced word processing before sufficient acoustic information had accrued to identify the word uniquely. The difference between congruous and incongruous sentence completions began not at the IP, but at the discrepancy point, when the acoustic input first diverged from expectations. For the cohort word pairs, these two time points are essentially the same,[5] and the difference between congruous and

[5] The point at which the words of a cohort pair diverged from each other was not evaluated in this study. For many of the stimuli, this is likely to be the same as the IP. For example, the auditory

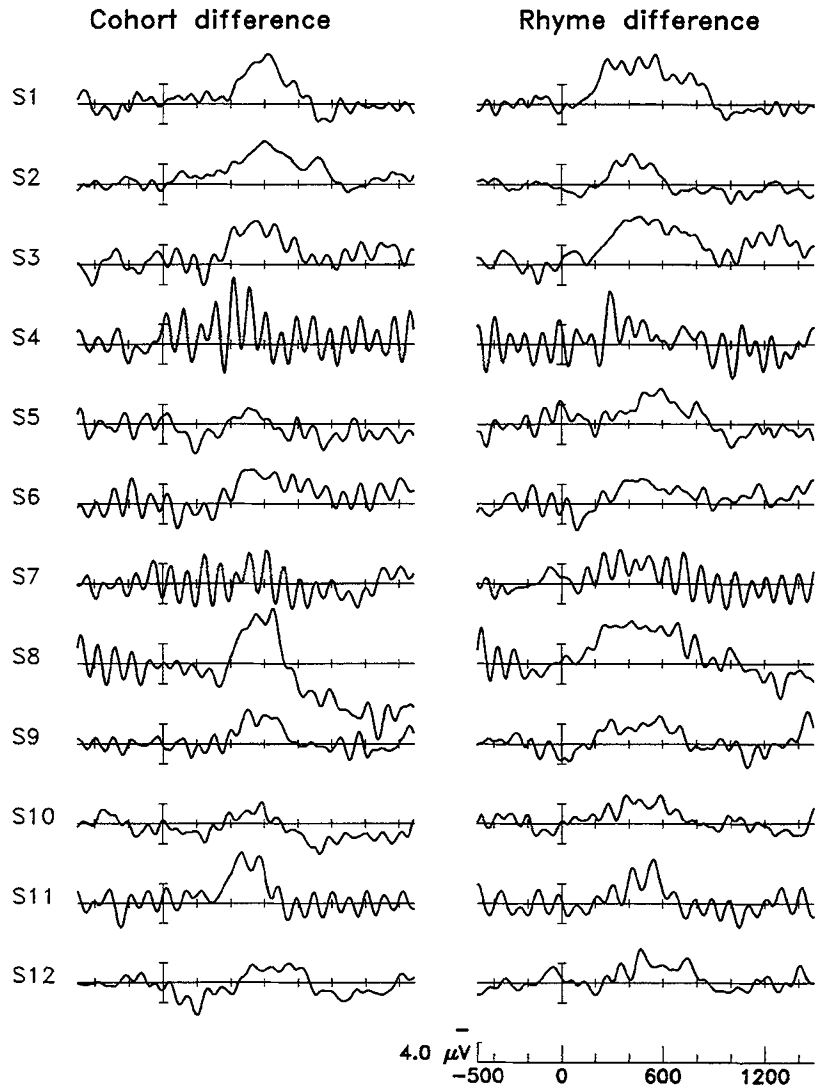

Figure 8. Difference waves from the first 12 individual participants, recorded at scalp site Pz. The left column displays the difference between the cohort incongruous and cohort congruous conditions. The right column displays the difference between the rhyme and cohort congruous conditions. S = subject.

incongruous sentence completions began very close to the IP. However, for the rhyming words, the discrepancy point occurs earlier, well in advance of the IP, and the N400 congruity also preceded the IP by some 200 ms (Figures 7 and 10). The results thus demonstrate an early link between phonological analyses and semantic integration.

input for DOLPHIN becomes unique at the third consonant, which is the same point that it diverges from its experimental associate DOLLAR. However, in other cases, it is clear that the separation between words of a pair must precede the IP for at least one of the words. For instance, DETAIL and DETOUR diverge at the second vowel, but DETAIL cannot uniquely be identified until its final consonant (given the existence of DETAIN).

Semantic or Phonemic Matching?

However, there are two accounts of this link, differing in their description of exactly what sort of expectation listeners derived from the sentence context. One possibility is that the semantic constraints of a sentence were used to form a phonemic template for a single word, and the N400 elicited by incongruous words reflected the mismatch with this *phonemic* expectation. The alternative account is that a *semantic* expectation was formed, and the early onset of the congruity effect reflects the initiation of semantic processes with only partial information about word identity. These accounts are difficult to distinguish because they share the core notion that incoming acoustic information is assessed

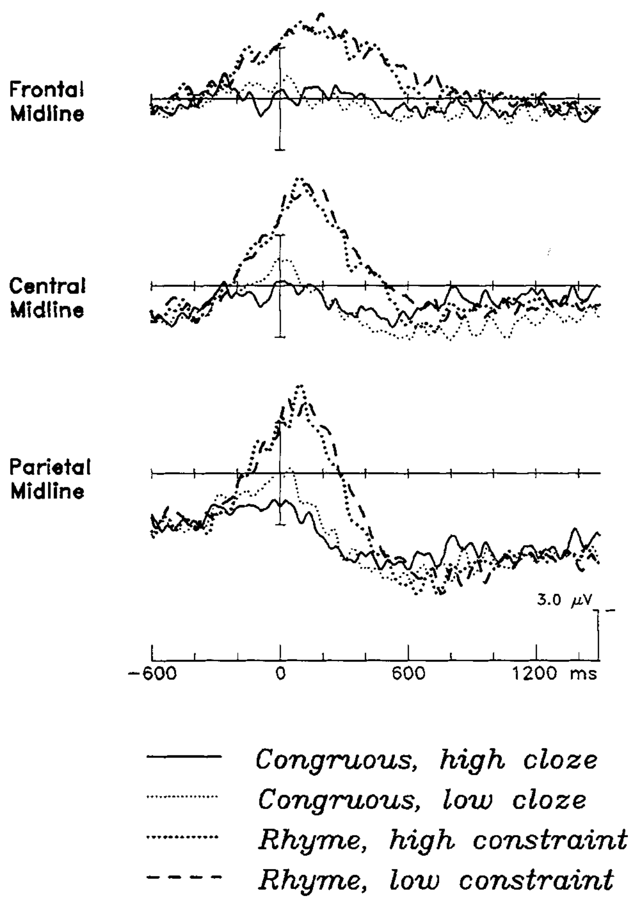

Figure 9. Grand average event-related potentials from the three midline scalp sites, timelocked to the isolation point (Time 0).

for its relationship to a preceding concept. The phonemic account stipulates that the semantic constraints of a sentence frame are translated into an expected phonological form for the upcoming word, and this phonemic template is then matched with the incoming acoustic signal. The semantic account includes no precomputation of possible phonological forms, but it instead assumes that incoming acoustic information receives preliminary semantic processing and that the results of these early semantic analyses are compared with the semantic constraints of the preceding sentence frame.

Although both accounts are compatible with our core finding that the sentence congruity effect can precede the IP, the balance of evidence argues against the phonemic-expectation account. Although the rhyming words were phonemically similar to the congruous endings, the ERPs elicited in the rhyme condition did not differ from those to dissimilar incongruous words (Figure 5, replicated in Experiment 3). The absence of any rhyming effect is informative with respect to the participants' processing mode. Other studies have shown that an N400-like potential reflects overall phonological similarity when the experimental conditions foster attention to such relationships. In auditory word lists with no semantic relationships and a high proportion of rhyming pairs, rhyming words elicit a smaller late negative wave than nonrhyming words presented for lexical decision (Praamstra, Meyer, & Levelt, 1994). In visual word lists, a similar rhyming effect is observed, but only if rhyme monitoring is the assigned task (Rugg & Barrett, 1987). The absence of a rhyme effect here suggests that participants were not engaged in the explicit detection of phonemic similarity or dissimilarity.

In contrast, numerous studies suggest that sentence stimuli engage semantic processing as the default strategy; semantic influences on N400 amplitude are observed independent of the overt behavioral task (Connolly et al., 1990; Kutas & Hillyard, 1980a; Kutas & Van Petten, 1994). Perhaps the most compelling demonstration that expectations based on sentence context are typically semantic is that incongruous words elicit smaller N400s when they are related to the congruous completion than when they are not (e.g., *UMBRELLA* vs. *DICTATOR* after a sentence frame such as "The game was called when it started to ___"; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). Because readers are unlikely to have generated a visual template for these wholly incongruous completions, this result argues that the N400 in the visual modality reflects semantic processing of the word presented, rather than a mismatch between the actual and expected visual forms.[6]

The analyses of cloze probability and contextual constraint in the current study suggest a similar conclusion for the auditory modality. For the eliciting words (those actually presented), N400 amplitude was dependent on the semantic goodness of fit with the preceding sentence frame: largest for wholly incongruous items, intermediate for low-cloze but acceptable items, and smallest for high-cloze endings (Figure 9). The difference between high- and low-cloze congruous items was of small amplitude, probably because the contrast was between two broad (and contiguous) ranges of probability rather than extreme values. In future research, it will be of some interest to examine several narrower ranges of probability. The small amplitude of the cloze effect did not allow a precise quantification of its onset latency, but like the overall congruity effect, it clearly began before the IP. The early onset of the cloze effect suggests that a dichotomous contrast between congruous and incongruous items is not an essential feature for the observation of early semantic processing but instead that anomalous words are only the endpoint of a continuum (see Kutas & Kluender, 1994, for an extended discussion of outright errors vs. less-favored continuations).

In contrast to the influence of cloze probability for congruous items, the degree of contextual constraint im-

[6] Additional evidence that the visual N400 is driven by semantic, rather than by perceptual, analyses of the eliciting word comes from studies that included visually surprising stimuli. Kutas and Hillyard (1980a, 1980c) presented sentence completions that were semantically acceptable, but in a much larger font than the sentence frames. These items elicited very different ERPs than acceptable items in the normal font, but the difference consisted of a late positive potential (P300) rather than an N400.

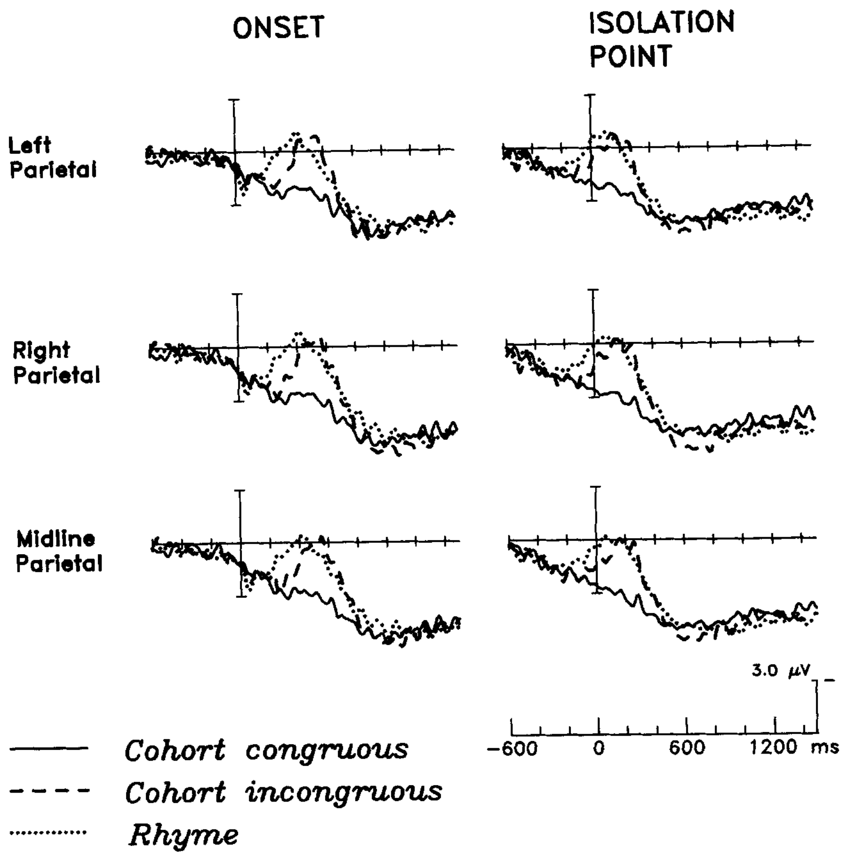

Figure 10. Grand average event-related potentials from Experiment 2, recorded from the three parietal scalp sites. In the left column, Time 0 is word onset; in the right column, Time 0 is the isolation point.

posed by the sentence frames had no influence on the amplitude or latency of the N400 elicited by the incongruous completions. This null result is contrary to the predictions of the phonemic-expectation account, which places heavy reliance on a listener's ability to predict a single sentence completion and to compute a phonemic template for that item. The phonemic account would suggest that if a listener does not generate a candidate completion, there will be no template to match or mismatch the incoming auditory signal and no N400 generated by the mismatch. The phonemic

Table 6
Time Course of the Sentence Congruity Effects in Experiment 3, Measured From Word Onset

Latency window (ms)	Cohort congruous vs. rhyme			Cohort congruous vs. cohort incongruous		
	$F(1, 17)$	*MSE*	$p<$	$F(1, 17)$	*MSE*	$p<$
0–100	0.85	8.67	*ns*	0.01	8.15	*ns*
100–200	0.26	10.98	*ns*	0.16	9.76	*ns*
200–300	28.35	6.98	.0001	0.82	6.96	*ns*
300–400	55.28	8.43	.0001	5.20	8.70	.05
400–500	47.56	12.69	.0001	31.64	10.65	.0001
500–600	28.38	14.18	.0001	68.06	8.30	.0001
600–700	15.88	17.04	.001	21.96	11.78	.0002
700–800	6.04	20.07	.05	1.99	24.97	*ns*
800–900	7.32	14.65	.02	0.08	30.56	*ns*
900–1,000	6.91	18.31	.02	0.13	26.62	*ns*

Note. Results of analyses of variance in which sentence type and electrode site were used as repeated measures: main effect of congruity. Latency window is in milliseconds after onset of sentence-final word.

Table 7
Time Course of the Sentence Congruity Effects in Experiment 3, Measured From Isolation Point

Latency window (ms)	Cohort congruous vs. rhyme			Cohort congruous vs. cohort incongruous		
	F(1, 17)	*MSE*	*p*<	*F*(1, 17)	*MSE*	*p*<
−500 to −400	0.03	2.9	*ns*	0.57	6.3	*ns*
−400 to −300	0.89	6.2	*ns*	0.01	7.8	*ns*
−300 to −200	0.73	4.8	*ns*	1.09	3.4	*ns*
−200 to −100	7.39	5.6	.02	0.01	5.0	*ns*
−100 to 0	7.77	9.1	.001	2.24	8.9	*ns*
0 to 100	51.70	10.1	.0001	15.50	10.2	.002
100 to 200	49.32	11.9	.0001	43.72	10.8	.0001
200 to 300	27.56	14.4	.0001	52.71	8.3	.0001
300 to 400	11.17	22.6	.005	6.85	17.3	.02
400 to 500	7.27	20.0	.02	0.64	30.5	*ns*
500 to 600	5.04	16.3	.05	0.05	24.5	*ns*
600 to 700	2.18	19.0	*ns*	0.03	20.7	*ns*
700 to 800	1.62	17.0	*ns*	0.03	13.3	*ns*

Note. Results of analyses of variance in which sentence type and electrode site were used as repeated measures: main effect of congruity. Latency window is in milliseconds before or after isolation point for sentence-final word (negative and positive numbers, respectively).

expectation account would thus have predicted that sentence frames without a single, strongly preferred completion should lead to small or delayed N400s as compared with sentence frames that do yield a strong candidate completion. This was not the case; instead, the results were similar to previous studies in the visual modality showing no influence of contextual constaint independent of cloze probability (Kutas & Hillyard, 1984; Kutas et al., 1984).

Overall, the pattern of results is most consistent with the conclusion that the listeners derived an expectation about word meaning from the sentence context and that the observed congruity effects reflected the semantic compatibility or incompatibility between this expectation and the meaning of the auditory input. The early onset latencies of the congruity effects then suggest that semantic integration begins with only partial sensory information about word identity. The similarities of the results across sentences varying in contextual constraint, and across the short and long interstimulus intervals used in Experiments 2 and 3, suggest that the early semantic integration observed here is a robust phenomenon.

Continuous Semantic Processing

How plausible is it to imagine that semantic processing can be initiated before the acoustic information specifies what word is being heard? The limits of parallel processing for the human listener are unknown, but it is difficult to imagine that the meaning of every word in one's vocabulary can be consulted at each moment in the speech signal. At least one empirical question and one theoretical question are thus relevant to this issue: How many words are still consistent with the input at the point when the congruity effect begins, and how we should conceptualize this early semantic processing. For the empirical question, it is important to note that the congruity effect for words with inappropriate initial phonemes did not begin at word onset, but rather some 150–200 ms into the word (Figures 4, 6, and 10). The gating data indicated that an average of two initial phonemes had been identified with 150 ms of input. One study reports that a dictionary count of American English yields a median set size of 87 words sharing two initial phonemes (Marslen-Wilson, 1984). A second dictionary count using duration rather than a phoneme measure reports that an average of 25 words remains possible with 150-ms signal durations and 11 with 200-ms durations (Wayland et al., 1989). These estimates show some variability, but both are likely to err on the large side given that not all of the dictionary words may be in any given listener's vocabulary. Neither estimate is so large as to rule out the possibility that a set of lexical candidates is activated by partial input and assessed in the context of the preceding sentence frame.

The theoretical question of the nature of early semantic processing is more difficult to approach given that our measures provide quantitative information about time course rather than qualitative information about the content of semantic processing. However, there is an interesting contrast between our two measures of word identification and semantic integration. The gating results include not only evidence of continuous use of acoustic input in word identification as shown by the progressive reduction in number of candidates across gates but also evidence of a genuine threshold when the input was adequate to select a single candidate. The electrophysiological measure of semantic integration was less suggestive of any discontinuities or sharp boundaries. Even when the ERPs were timelocked to the IP—the apparent threshold for word identification—none of the observed semantic context effects were discrete but instead extended over several hundreds of milliseconds. Moreover, the pattern of the latency results did not suggest that the N400 congruity effect was an all-or-none process triggered at variable time points. For words with inappropriate initial phonemes (both the rhyme and plain incongruous conditions), the congruity effect began by 200 ms after word onset and continued to about 1,000 ms. For words with consistent initial phonemes (cohort incongruous), the congruity effect was delayed in onset until some 400 ms but was also over and done by 1,000 ms. In other words, N400 amplitude remained low as long as the acoustic input was consistent with the sentence context but increased rapidly when the auditory signal became inconsistent with a congruous sentence completion. It is possible to think of the cohort incongruous condition as encouraging a semantic "garden path"—the auditory input is initially consistent with a plausible sentence completion but then turns into an anomalous word. However, in contrast to other garden-path phenomena, there was little sign that the first conflicting input (at the IP) triggered new processes of repair or reanalysis. Instead, the ERPs elicited by the cohort incongruous items rapidly grew to resemble the ERPs elicited by the other incongruous items, which were never consistent with the sentence frame. The latency results are more compatible with a continuous mapping from acoustic input onto semantic representations, rather than a single selection point at which an item is classified as congruous or incongruous and semantic processing must be restarted if the wrong choice is made.

Overall, the results suggest that semantic integration during speech comprehension is a temporally extended process which begins before word identification is complete, which also continues for some time afterward. The findings are most compatible with general models of information processing that allow transmission of partial information between temporally overlapping stages of analysis (J. O. Miller, 1993).

References

Altmann, G. T. M. (Ed.). (1990). *Cognitive models of speech processing: Psycholinguistic and computational models.* Cambridge, MA: MIT Press.

Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1278–1296.

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause—Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology, 46,* 3–23.

Bradley, D. C., & Forster, K. I. (1987). A reader's view of listening. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 103–134). Cambridge, MA: MIT Press.

Brown, C. M. (1990). *Spoken-word processing in context.* Doctoral dissertation, Katholieke Universiteit Nijmegen, the Netherlands.

Chwilla, D. J. (1996). *Electrophysiology of word processing: The lexical processing nature of the N400 priming effect.* Doctoral dissertation, Katholieke Universiteit Nijmegen, the Netherlands.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26,* 527–538.

Connine, C. M. (1990). Effects of sentence context and the lexicon during speech processing. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 281–294). Cambridge, MA: MIT Press.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193–210.

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 & Psychophysics, 56,* 624–636.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6,* 256–266.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related potentials. *Electroencephalography and Clinical Neurophysiology, 94,* 276–287.

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

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39,* 302–318.

Craig, C., Kim, B., Rhyner, P., & Chirillo, T. (1993). Effects of word predictability, child development and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research, 36,* 832–841.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology, 21,* 105–113.

Dechert, H. W., & Raupach, M. (Eds.). (1980). *Temporal variables in speech: Studies in honor of Frieda Goldman-Eisler.* The Hague, the Netherlands: Mouton.

Deese, J. (1984). *Thought into speech: The psychology of a language.* Englewood Cliffs, NJ: Prentice Hall.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Ford, J. M., Woodward, S. H., Sullivan, E. V., Isaacks, B. G., Tinkleberg, J. R., Yesavage, J. A., & Roth, W. T. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography and Clinical Neurophysiology, 99,* 235–246.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*(A), 465–495.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Frauenfelder, U. H., & Tyler, L. K. (Eds.). (1987). *Spoken word recognition.* Cambridge, MA: MIT Press.

Garrett, M. F. (1990). Sentence processing. In D. N. Osherson & H. Lasnik (Eds.), *An invitation to cognitive science: Language* (pp. 133–176). Cambridge, MA: MIT Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15,* 103–113.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28,* 267–283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics, 38,* 299–310.

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

Jasper, H. H. (1958). The ten–twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

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

Kutas, M., & Hillyard, S. A. (1980b). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11,* 99–116.

Kutas, M., & Hillyard, S. A. (1980c). Reading between the lines: Event related brain potentials during natural sentence processing. *Brain and Language, 11,* 354–373.

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

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience, 1,* 38–49.

Kutas, M., & Kluender, R. (1994). What is who violating? A reconsideration of linguistic violations in light of event-related brain potentials. In H.-J. Heinze, T. F. Munte, & G. R. Mangun (Eds.), *Cognitive electrophysiology* (pp. 183–210). Boston: Birkhauser.

Kutas, M., Lindamood, T., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Komblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–238). Hillsdale, NJ: Erlbaum.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 3, pp. 139–187). Greenwich, CT: JAI Press.

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

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–302). San Diego, CA: Academic Press.

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, MA: MIT Press.

Marslen-Wilson, W. (1984). Function and process in spoken word-recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention & performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71–102). Cambridge MA: MIT Press.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187–210). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–1392.

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

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576–585.

Massaro, D. W. (1994). Psychological aspects of speech perception: Implications for research and theory. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 219–264). San Diego, CA: Academic Press.

McAllister, J. (1988). The use of context in auditory word recognition. *Perception & Psychophysics, 44,* 94–97.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 59,* 477–488.

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

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Journal of Language and Cognitive Processes, 10,* 309–332.

Miller, J. L., Green, K., & Schermer, T. (1984). A distinction between effects of sentential speaking rate and semantic congruity on word recognition. *Perception & Psychophysics, 36,* 327–336.

Miller, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review, 100,* 702–715.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Moss, H. E., & Marslen-Wilson, W. (1993). Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1254–1276.

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). San Diego, CA: Academic Press.

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93–136.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

O'Rourke, T. B., & Holcomb, P. J. (in press). Electrophysiological evidence for the efficiency of spoken word processing. *Language and Cognitive Processes.*

Praamstra, P., Meyer, A. S., & Levelt, W. J. M. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience, 6,* 204–219.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Regan, D. (1989). *Human brain electrophysiology.* New York: Elsevier.

Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme–judgement task. *Brain and Language, 32,* 336–361.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24,* 210–231.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. L. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315–328.

Sereno, S. C. (1995). Resolution of lexical ambiguity: Evidence from an eye movement priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 582–595.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359–374). New York: Academic Press.

Stirling, L., & Wales, R. (1996). Does prosody support or direct sentence processing? *Language and Cognitive Processes, 11,* 193–212.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science, 155,* 1436–1439.

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.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics, 36,* 417–427.

Tyler, L. K. (1990). The relationship between sentential context and sensory input: Comments on Connine's and Samuel's chapters. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 315–323). Cambridge, MA: MIT Press.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word recognition processes. *Perception & Psychophysics, 34,* 409–420.

Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics, 38,* 217–222.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes, 8,* 485–532.

Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology, 32,* 511–525.

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language, 26,* 188–208.

Van Petten, C., & Kutas, M. (1988). The use of event-related potentials in the study of brain asymmetries. *International Journal of Neuroscience, 39,* 91–99.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18,* 380–393.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory & Cognition, 19,* 95–112.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8,* 238–242.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41,* 262–275.

Wayland, S., Wingfield, A., & Goodglass, H. (1989). Recognition of isolated words: The dynamics of cohort reduction. *Applied Psycholinguistics, 10,* 475–487.

Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology, 87,* 306–320.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken word processing. *Cognition, 32,* 25–64.

Received July 14, 1997
Revision received July 20, 1998
Accepted July 24, 1998 ■

ing: Psycholinguistic and computational perspectives (pp. 315–323). Cambridge, MA: MIT Press.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word recognition processes. *Perception & Psychophysics, 34,* 409–420.

Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics, 38,* 217–222.

semantic context in spoken word processing. *Cognition, 32,* 25–64.

Received July 14, 1997
Revision received July 20, 1998
Accepted July 24, 1998 ■