The Cascaded Nature of Lexical Selection and Integration in Auditory Sentence Processing

Daniëlle van den Brink, Colin M. Brown, and Peter Hagoort
F. C. Donders Centre for Cognitive Neuroimaging

An event-related brain potential experiment was carried out to investigate the temporal relationship between lexical selection and the semantic integration in auditory sentence processing. Participants were presented with spoken sentences that ended with a word that was either semantically congruent or anomalous. Information about the moment in which a sentence-final word could uniquely be identified, its isolation point (IP), was compared with the onset of the elicited N400 congruity effect, reflecting semantic integration processing. The results revealed that the onset of the N400 effect occurred prior to the IP of the sentence-final words. Moreover, the factor early or late IP did not affect the onset of the N400. These findings indicate that lexical selection and semantic integration are cascading processes, in that semantic integration processing can start before the acoustic information allows the selection of a unique candidate and seems to be attempted in parallel for multiple candidates that are still compatible with the bottom–up acoustic input.

Keywords: event-related potentials, N400, auditory sentence processing, semantics, lexical selection

Understanding spoken language happens seemingly effortlessly. However, in reality, it is no simple exercise. On the contrary, when people hear speech, numerous brain areas work together to analyze the acoustic information, select the proper words by mapping the sensory input onto stored lexical knowledge, extract the meanings of those words, and integrate them into an ongoing sentential or discourse context (cf. Marslen-Wilson & Welsh, 1978; Norris, 1994; Zwitserlood, 1989). All of this happens in half a second. In this study, we address the question of whether lexical selection and semantic integration of a word's meaning into the preceding context are two sequential processes that can be separated in time or whether they, to a certain extent, run in parallel.

Models of spoken word processing converge on the idea that upon hearing the initial phoneme(s) of a word, multiple lexical candidates that match the acoustic input are activated in parallel (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; McQueen, Norris, & Cutler, 1994; Norris, 1994; Norris, McQueen, & Cutler, 1995). During the lexical selection phase, these candidates are assessed for goodness of fit within the preceding sentence or discourse context. In the presence of candidates in the activated set that fit the sentence context well, further incoming acoustic information and top–down contextual information are used to narrow down the number of candidates to the one that is most compatible with both form and

content constraints. This candidate is integrated in the sentence context. If, however, none of the lexical candidates fit the context well, selection of the proper candidate is difficult and can only be done on the basis of the acoustic information. We are interested in whether there is a discrete moment in which lexical selection ends and semantic integration begins or whether these two processes are of a cascading nature with the semantic integration process initiating before lexical selection can be completed.

To probe the (end of the) selection phase, we determined the moment at which a word can be uniquely identified, defined as the minimum amount of acoustic signal necessary to identify a spoken word in the absence of contextual information. This moment is referred to as the isolation point (IP) of a word. IPs can be obtained with the gating technique (Grosjean, 1980). In the gating paradigm, participants hear increasingly longer fragments of words (e.g., in steps of 50 ms), and they are asked for each fragment or *gate* to guess what word is being presented and to indicate how confident they are about their guess. Typically, when presented with the first gates, participants generate many different candidate words while being quite unsure about their guesses. However, as the length of the acoustic signal is increased, they tend to converge on one candidate, and their confidence becomes very high. The IP is defined as the total duration of the first gate at which most participants generate the correct word and do not change their opinion at successive gates. The confidence rating is not included in the IP measure but is incorporated in a related measure, the recognition point (RP) measure. The RP is defined as the gate at which participants generate the correct word and are highly confident about their guess. The IP either coincides with or precedes the RP and is therefore taken to be the earliest moment at which a word can be identified.

A number of behavioral studies have shown that, in the presence of a semantically constraining sentence context, the IP is moved forward, that is, the duration of the signal that is necessary to

identify the word is shortened, and the number of possible word candidates is reduced (Grosjean, 1980; Salasoo & Pisoni, 1985; Tyler & Wessels, 1983; Zwitserlood, 1989). These findings indicate that selection of the proper word is aided by the contextual constraints. However, in the case of a semantically anomalous word, the context cannot aid in selecting a proper word, and therefore IP remains the same. In these instances, is semantic integration postponed until one candidate has been selected on the basis of acoustic input alone (i.e., at IP), or is integration attempted for more than one word candidate that still matches the acoustic input?

A technique that seems especially suitable for studying brain processes in real time is the recording of event-related potentials (ERPs). ERPs reflect the sum of simultaneous postsynaptic activity of a large population of neurons recorded at the scalp as small voltage fluctuations in the electroencephalogram. Because of the high temporal resolution of this technique, it has proven to be a sensitive measure of real-time language processing. A well-documented finding in the study of language processing is the N400 component, a negative polarity potential that typically starts at about 200–250 ms after word onset, peaks at 400 ms, and has a centro-parietal distribution across the scalp (cf. Kutas & Hillyard, 1980). Previous ERP studies have revealed that the N400 component is related to semantic processing of the eliciting word. More specifically, the N400 amplitude indexes the relative ease with which the meaning of a word is integrated in the context. The better the semantic fit of a word within the preceding sentence frame or larger discourse, the more reduced the amplitude of the N400 (Brown & Hagoort, 1993; Hagoort & Brown, 2000; Holcomb, 1993; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003).

In the current study, we evaluated the temporal relationship between lexical selection and semantic integration by means of comparing the IP of a word with the onset of the N400 congruity effect in the ERP signal. To that end, we presented participants with sentences that ended with a word that was either semantically congruent or anomalous (e.g., "This 14-year-old pianist obviously has a lot of *talent*" or "This 14-year-old pianist obviously has a lot of *climate*"). IPs for all critical words were assessed by means of the gating paradigm. This allowed us to align the electroencephalogram (EEG) waveforms of the individual sentence completions to the IP for each word to evaluate whether IP correlated with the onset of the N400 effect. Furthermore, we divided the critical words into two groups with early or late IPs, resulting in four experimental conditions: (a) congruent completions with early IPs, (b) incongruent completions with early IPs, (c) congruent completions with late IPs, and (d) incongruent completions with late IPs. If semantic integration into the prevailing context does not set in before the lexical selection process is completed, then the N400 effect measured between words that fit a sentence context and words that do not fit a sentence context should set in at or soon after IP and should therefore differ between the early and late IP words. If, however, lexical selection and semantic integration are cascading processes, and semantic integration works with several candidates, then onset of the N400 could occur prior to the IP and would not have to differ as a function of IP latency.

In a recent ERP study, Van Petten, Coulson, Rubin, Plante, and Parks (1999) have also used the sensitivity of the N400 and IPs of words to examine the temporal relationship between the initiation of semantic integration processing and the moment that a word can

be identified uniquely. Their experimental items consisted of sentences that ended either with (a) the highest cloze probability word, (b) a word that rhymed with the highest cloze probability word, (c) a semantically anomalous word that shared the same initial phonemes as the highest cloze probability word, or (d) an anomalous word of which the initial phonemes differed from those of the highest cloze probability word. For example, "It was a pleasant surprise to find that the car repair bill was only 17 *dollars/ scholars/dolphins/bureaus*." The results revealed that the sentence-final words in the three anomalous conditions all elicited a significantly larger N400 than the high-cloze probability endings. Moreover, the onset of the N400 effect differed between conditions. In the fully anomalous and rhyme conditions, the onset of the N400 effect preceded the IP of the sentence-final words. Relative to these two conditions, the onset of the N400 effect in the condition in which the anomalous word shared its phonemic onset with the highest cloze probability word was found to be delayed by some 200 ms and corresponded to the IP. Van Petten et al. (1999) concluded that the N400 effect reflects semantic integration processing of the perceived word, and that integration works with partial and incomplete information about the perceived word.

Note that the stimulus materials used in Van Petten et al.'s (1999) study did not consist of continuous speech. Rather, the sentence frames were spoken and recorded without their final words, and the critical words were also spoken in isolation. When presenting the sentences to the participants, the critical words were spliced onto the sentence frame and were preceded by a period of silence. The duration of silence preceding the sentence-final word was manipulated in two experiments (500 ms vs. 50 ms) and led to similar results, indicating that a guessing strategy due to the unnatural long pause of 500 ms before the sentence-final word was not a possible factor for the outcomes. However, in both experiments, the combination of recording and splicing resulted in unnatural intonation patterns of the materials, which still might have affected the results. In the current experiment, we set out to replicate the above mentioned findings made by Van Petten et al. involving the congruent and semantically anomalous sentence completions, using natural connected speech. Moreover, we added an additional IP manipulation by splitting the critical words into two groups with early and late IPs, resulting in clear predictions about the possible relation between IP and the onset of the N400 effect.

Method

Participants

The experiment was conducted with 12 native speakers of Dutch (11 women) with a mean age of 21.5 years (range = 19–23 years) from the subject pool of the Max Planck Institute for Psycholinguistics (Nijmegen, the Netherlands). All had normal to corrected-to-normal vision and were right-handed (3 participants reported having a left-handed parent or sibling). None of the participants had any neurological impairment or had experienced any neurological trauma. In addition, none of the participants participated in the pretest and gating study (see below). Participants were paid for their participation.

Materials

The experimental items consisted of a set of 261 semantically constraining Dutch sentences with terminal words that varied across two experi-

mental conditions. The sentences were identical up to the final word. The sentence-final word was either congruent or anomalous. In the congruent condition, sentences ended with the highest cloze probability word, for example, "Deze veertienjarige pianist heeft duidelijk veel *talent*" ("This 14-year-old pianist obviously has a lot of *talent*"). In the incongruent condition, sentences ended with a word that was semantically anomalous and also had initial phonemes that differed from the highest cloze probability word, for example, "Deze veertienjarige pianist heeft duidelijk veel *klimaat*" ("This 14-year-old pianist obviously has a lot of *climate*").

The 261 constraining sentences were derived from a set of sentence materials that had been used in a previous study (van den Brink, Brown, & Hagoort, 2001) and which were part of a larger set of sentences that had been submitted to a cloze test. Thirty students from the Radboud University (Nijmegen, the Netherlands) were given a list of 350 sentences with the final word omitted. Participants were asked to carefully read the sentences and to fill in the first word that came to mind at the position of the omitted sentence-final word. On the basis of the results, a set of 261 sentences with a cloze probability of .50 or higher formed the basis for the materials used in this experiment. Items for the incongruent condition were constructed by replacing the sentence-final word with a semantically anomalous word with word-initial phonemes that differed from the beginning phonemes of the high cloze probability word and other congruent words.

For purposes not relevant to this study, we added a third condition also containing semantically anomalous sentence-final words. This third condition was not analyzed as part of this experiment. The experimental materials consisted of a set of 261 semantically constraining sentences with sentence-final words that varied across three experimental conditions (with 87 items per condition per participant). Two of these conditions contributed data to the analyses. Because the sentences were identical up to the final word, they had the same length (mean length of 10.8 words; range = 5–15 words) and were equally constraining; the critical words in the congruent condition had a mean cloze probability of .84 (range = .50–1.0). Because of inclusion of the third anomalous condition, constraints on the construction of the experimental materials demanded that we tested between critical items. We, therefore, matched the critical items on several aspects. All sentence-final words were reasonably well-known nouns selected from a Dutch word corpus called CELEX (Baayen, Piepenbrock, & Van Rijn, 1993); all critical words had a frequency of at least 30 counts per 42 million (this equals 1.48 log), with a mean duration of 516 ms and a mean log frequency of 2.87 in the congruent condition, and 515 ms and 2.49 in the incongruent condition. Each sentence-final word began with the plosive /p/, /t/, /k/, /b/, or /d/. This provided a clear physical marker that facilitated the alignment of the ERP waveform to the onset of the sentence-final word.[1] To ensure that any obtained congruency effect could not be the result of an item effect due to the use of different plosives as onset phonemes, we matched the distribution of onset phonemes across the congruent and anomalous conditions: Number of voiced plosives as onset phonemes was 105 in the congruent condition versus 108 in the incongruent condition, and number of voiceless plosives as onset phonemes was 155 in the congruent condition versus 152 in the incongruent condition.

Three lists were constructed to ensure that no participant heard the same sentence or critical word more than once. In addition to the set of 261 experimental items (87 per condition per participant), a set of 87 filler items was constructed to balance the number of sentences that ended anomalously and congruently. Finally, the experimental lists were preceded by a practice list of 20 items, which reflected the experimental materials. The practice list was used to familiarize the participants with the experimental procedure.

The experimental sentences, carrier phrases, fillers, and practice sentences were all spoken by a female speaker, with normal intonation at a normal speaking rate, and they were recorded in one recording session. The sentences were recorded in triplets (3 times the same sentence frame with a different sentence-final word; see van den Brink et al., 2001) to make sure that intonation and speaking rate were kept constant within each set.

Sentences were spoken in a sound-attenuating booth and recorded on a digital audiotape. The digital audio recordings were sampled at 16 kHz mono and stored on computer disk. A speech waveform editing system was used to mark the onset of the sentence-final words.

Procedure

Participants were tested individually in a dimly illuminated sound-attenuating booth. They were seated in a comfortable reclining chair, instructed to move as little as possible, and told that they would hear a number of sentences. Their task was to attentively listen to these sentences and to try to understand them. In addition, they were told that several sentences would be semantically anomalous. No additional task demands were imposed.

Each trial began with a 300 ms warning tone, followed after by 1,200 ms of silence and then by the spoken sentence. The next trial began 4,100 ms after sentence offset. To ensure that participants would not blink during and shortly after presentation of the sentence, we displayed a fixation point (an asterisk) on the computer screen 1,000 ms prior to the beginning of the sentence. The asterisk remained on the screen until 1,600 ms after offset of the spoken sentence. Participants were instructed to fixate on the asterisk during presentation of the sentences, but they were free to blink and move their eyes when the asterisk was not displayed on the screen. Participants listened to the stimuli via closed-ear headphones. After the practice session, the experimental trials were presented in five blocks of approximately 10 min.

EEG Recording

The EEG was recorded from 29 Ag/AgCl-sintered electrodes mounted in an elastic cap, each referred to the left mastoid. Twenty-three electrodes were placed according to the 10–10 Standard System of the American Electroencephalographic Society (Fz, FCz, Cz, Pz, Oz, AF3, AF4, F3, F4, F7, F8, FC3, FC4, FT7, FT8, C3, C4, CP3, CP4, P3, P4, PO7, and PO8). Three additional pairs were placed laterally over symmetrical positions: (a) a temporal pair (LT and RT) placed laterally to Cz, at 33% of the interaural distance, (b) a temporo-parietal pair (LTP and RTP) placed 30% of the interaural distance lateral and 13% of the nasion-inion distance posterior to Cz, and (c) a parietal pair midway between LTP/RTP and PO7/PO8 (LP and RP). Vertical eye movements were monitored via a supra- to suborbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye movements. Activity over the right mastoid

---

[1] In the study of phonetics, the voiceless plosive is assumed to consist of the following three phases: (a) the catch, when the airway closes so that no air can escape through the mouth; (b) a period of silence, called *closure*, when the articulators totally block the air stream—during this period of closure the air pressure increases just behind the obstruction; and (c) an audible *burst* of air, when the articulators quickly move away from each other. In the English language, but not in Dutch, the burst of the voiceless plosive is followed by aspiration, when the articulators are further apart, and the air pressure at the site of the obstruction has fallen so that the speech sound is no longer a burst but is still audible. Voiced plosives also consist of three phases. However, during Phase 2, the period of closure, the vocal cords are vibrating. One might argue that in running speech, critical words starting with voiceless plosives should have an onset marker placed at the burst, because this is the first audible aspect of the phoneme. However, because the period of closure is considered to be an intricate part of the plosive, we treated the "period of silence" preceding the burst as a first cue to the onset phoneme of the critical word. Therefore, we consistently marked the *onset of closure*, that is, the "silence" preceding the burst in voiceless plosives, and the onset of vocal cord vibration before the burst in voiced plosives, as the onset of the critical words.

bone was recorded on an additional channel to determine whether there were differential contributions of the experimental variables to the presumably neutral mastoid site. No such differential effects were observed.

We amplified the EEG and electro-oculogram (EOG) recordings with a SynAmp Model 5083 EEG amplifier (NeuroScan, Herndon, Virginia) using a high cut of 30 Hz (notch filter 60 Hz) and a time constant of 8 s (0.02 Hz). Electrode impedances were kept below 3 kOhm for the EEG recording and below 5 kOhm for the EOG recording. The EEG and EOG signals were digitized online with a sampling frequency of 200 Hz.

Data Analysis

Data from critical trials were analyzed according to the following procedure. Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking, and electromyogram artifacts in a critical window that ranged from 150 ms before onset of the sentence-final word to 1,900 ms after onset of the sentence-final word. Trials containing artifacts were rejected (average number of trials per condition = 6.5%, range = 4.1%–8.1%). For each participant, average waveforms were computed across all remaining trials per condition after applying a baseline correction to the waveforms of the individual trials on the basis of the averaged activity in a time window of 150 ms preceding onset of the sentence-final word.

On the basis of the results of a gating study (see below), we divided the critical words into equally large groups with early and late IPs, resulting in four conditions: congruent completions with early IP, incongruent completions with early IP, congruent completions with late IP, and incongruent completions with late IP. We performed statistical analyses of the congruity effects, which consisted of repeated measures analyses of variance (ANOVAs) with mean amplitude values computed for each participant and each electrode in the N400 latency window: 300–500 ms after final-word onset. The results were analyzed in an omnibus ANOVA with two levels of the IP factor (Early, Late), two levels of the congruity factor (Congruent, Anomalous), and the 29-level electrode factor. In addition to the omnibus ANOVA, we tested a priori pairwise comparisons between the congruity conditions using ANOVAs with a two-level congruity factor. Scalp distributions of the congruity effects were subsequently explored in three separate ANOVAs: an ANOVA on anterior versus posterior sites (Fz, FCz, AF3, AF4, F3, F4, F7, F8, FC3, FC4, FT7, FT8/Pz, Oz, CP3, CP4, LTP, RTP, P3, P4, LP, RP, PO7, PO8), anterior left hemisphere versus anterior right hemisphere (AF3, F3, F7, FC3, FT7/AF4, F4, F8, FC4, FT8), and posterior left hemisphere versus posterior right hemisphere (CP3, LTP, P3, LP, PO7/CP4, RTP, P4, RP, PO8). Univariate F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse–Geisser/Box's epsilon hat correction (Maxwell & Delaney, 1990). The original degrees of freedom and adjusted p values are reported below.

Results

Gating Experiment

To obtain the IP for each sentence-final word in the congruent and incongruent conditions, we had an additional group of 24 students perform an extensive gating task on the 522 isolated sentence-final words (261 sentences each with 2 completions; congruent and incongruent). On each trial, participants were auditorily presented with increasingly longer fragments (in steps of 50 ms) of one particular word. Amplitude over the final 10 ms of the fragments was tapered to zero to avoid clicks in the signal that would result from a sharp cutoff. For each fragment, the participants were asked to type in on a keyboard what they thought the word was going to be and, in addition, to give a confidence rating.

IP of a word was defined as the gate at which at least 67% of the participants correctly identified the word, and accuracy remained above 60% at subsequent gates.[2] The RPs of words were also measured for comparison purposes. RP of a word was defined as the gate at which all 24 participants correctly identified the word with 90% confidence, and accuracy remained at 100% at successive gates.

Table 1 displays the results of the gating experiment. On the basis of the results, we divided the critical words into equally large groups with early and late IPs. For the congruent completions, the 50th percentile corresponded to an IP at 283 ms and for the anomalous completions at 285 ms. This resulted in the following four conditions (see Table 2): congruent and incongruent completions with early IPs (on average about 230 ms), and congruent and incongruent completions with late IPs (on average about 340 ms).

ERP Experiment

Figure 1 displays the grand average waveforms of all four conditions by electrode site time-locked to the onset of the sentence-final word. What can be seen from the figure is that irrespective of whether a critical word belongs to the group with early or late IPs, the waveforms of the congruent and anomalous completions across these groups are practically identical. In both conditions, the sentence-final words elicited an N100 component. This component is most visible in the frontal to central electrodes, including the temporal electrodes. It is not common to find a clear N100 component in a study that has used natural connected speech, given the continuous nature of the speech signal (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990). However, as in previous studies of ours (van den Brink et al., 2001; van den Brink & Hagoort, 2004), all sentence-final words began with a plosive, which is a relatively clear onset marker in connected speech. In addition, the homogeneity of the sentence-final word onsets presumably added to the clear presence of the N100 component in the averaged waveforms. Furthermore, a recent ERP study of continuous speech processing by Sanders and Neville (2003), investigating a possible word-onset effect, also revealed the presence of an N100 to word-initial sounds in continuous speech.

Also, a negativity at approximately 200 ms is visible in the waveforms. This negativity is apparent in all four conditions but is largest in the incongruent conditions and is most visible over the frontal sites. In previous studies of ours (van den Brink et al., 2001; van den Brink & Hagoort, 2004), we have proposed that the N200 effect is indicative of the lexical selection process, in which word-form information (resulting from an initial phonological analysis) and content information (derived from the context) first interact. Because the current experiment was designed to investigate the possible correlation between IP and initiation of semantic integration, our primary interest concerns the broad negativity, apparent in the anomalous conditions peaking at approximately 400 ms, which are more negative than the ERPs elicited in the

---

[2] These criteria are similar to those used in a study by Van Petten et al. (1999), who defined IP as the gate at which 70% of the participants (7 of 10) correctly identified the word, and accuracy remained above 60% at subsequent gates.

Table 1
Results of the Gating Study Involving Critical Words Used in the ERP Experiment

Condition	IP	RP	Word duration
Congruent	282 (73)	383 (58)	516 (99)
Incongruent	290 (68)	380 (56)	515 (88)

Note. Means, with standard deviations in parentheses, are presented in milliseconds. $n = 261$ per condition. ERP = event-related potential; IP = isolation point: first gate at which 67% of 24 participants identified word correctly; RP = recognition point: first gate at which all 24 participants identified word correctly with at least 90% confidence.

congruent conditions. Its latency characteristics and morphology are similar to previously reported N400 effects. Statistical analysis on the mean amplitude values in the N400 latency interval revealed a significant main effect of congruity, $F(1, 11) = 57.62$, $p < .001$, $MSE = 88.97$, no main effect of IP, $F(1, 11) = 1.57$, $p = .236$, $MSE = 32.46$, and no interaction between these two factors ($F < 1.00$).

Table 3 displays the results of the a priori pairwise comparisons in the 300–500 ms latency range. Statistical analyses in this latency interval revealed that both incongruent completions elicited a larger N400 than the congruent completions (an effect of 3.81 μV for the early IP contrast and 3.86 μV for the late IP contrast). Interactions with electrodes were also obtained in the pairwise comparisons. Topographical analyses showed that both congruity effects in relation to the congruent conditions were significantly larger over posterior than anterior regions of the scalp. No differences between hemispheres were found, indicating that both N400 effects have classical centro-parietal distributions. Figure 2 shows the topography of mean amplitude effects between the congruent and incongruent conditions for the early and late IP contrasts in the N400 latency window. The figure underscores the statistical results, revealing that both N400 effects have centro-parietal distributions.

In Figure 1, ERPs were time locked to final-word onset. Figure 3 displays the difference waveforms of both early and late comparisons by electrode site time locked to IP, thereby correcting for variability of IP between individual words. The latency interval preceding IP corresponds to the mean duration preceding IP for all critical words per experimental contrast. Line E marks the mean duration preceding IP for the congruent and incongruent critical words with early IP, which is 229 ms. Line L marks the mean duration preceding IP for the congruent and incongruent critical words with late IP, which is 343 ms. Because information about the exact onset of each critical word is lost in this time-locking procedure, the zero point only roughly corresponds to word onset in Figure 3. Baseline correction of the waveforms was applied on the basis of the averaged activity in a time window of 150 ms preceding mean onset of the critical word (379–229 ms before IP for early contrast and 493–343 ms before IP for late contrast, corresponding to -150 to 0 ms in the figure). Again, what we see is that irrespective of which group the words belong to, the waveforms are remarkably similar, and a congruity effect is initiated before IP.

Onset analyses were performed on the IP time-locked data to establish the onset of the congruity effects. Onset latencies were

estimated by first separately computing the mean amplitude values for all 29 electrodes in 20-ms latency ranges (bins) that shifted in steps of 10 ms from 200 ms prior to IP until 100 ms after IP (e.g., -200 ms until -180 ms; -190 ms until -170 ms). The values for the latency bins were submitted to ANOVAs that tested against the null-hypothesis of zero difference between the a priori selected congruent versus incongruent conditions for words with early and late IPs separately. The onset latency analyses for congruent versus incongruent completions with early IP revealed a significant congruity effect in the -50 to -30 ms bin and in all following bins. The onset latency analyses for the congruent versus incongruent late IP comparison revealed a significant congruity effect in the -140 to -120 ms bin and in all following bins. So, in the case of the words with early IPs, mean IP is 230 ms, and the congruity effect sets in approximately 50 ms prior to IP. In the case of the words with late IPs, mean IP is 340 ms, and the congruity effect sets in approximately 140 ms prior to IP.

Discussion

The current study was designed to investigate the temporal relationship between lexical selection and semantic integration in auditory sentence processing. We investigated whether there is a discrete moment in which lexical selection ends and semantic integration begins or whether these two processes are of a cascading nature with semantic integration processing starting before lexical selection is completed. Information about the earliest moment at which a word can be identified, the IP of words, was used to investigate the onset of the N400 effect, an index of semantic integration processing. A gating experiment of 522 spoken words revealed that whereas the duration of words was on average 516 ms, mean IP was well before word offset, namely at 286 ms. These words served as congruent and incongruent completions of spoken sentences. The results of the gating experiment were used to divide the critical words into two groups. One group contained congruent and incongruent sentence-final words with early IPs (with a mean of 230 ms), and the other group consisted of congruent and incongruent completions with late IPs (mean of 340 ms). We hypothesized that if semantic integration processing is dependent on word identification, then the N400 effect should set in at or soon after IP and should therefore differ between the early and late IP words. If, however, semantic integration starts before the word is uniquely recognized, then onset of the N400 could occur prior to the IP and should not have to differ as a function of IP latency. The results revealed that, despite a mean difference in IP of at least 100 ms between the two groups of items, the factor early or late IP did not affect the onset, nor the peak, of the N400. Incongruent completions in both groups elicited an N400 before IP. This indicates that irrespective of whether the selection process had

Table 2
Mean IP Values for Critical Words in Four Conditions

Condition	Early IP	Late IP
Congruent	223 (40)	341 (46)
Incongruent	235 (38)	344 (44)

Note. Means, with standard deviations in parentheses, are presented in milliseconds. $n = 87$ per condition. IP = isolation point.
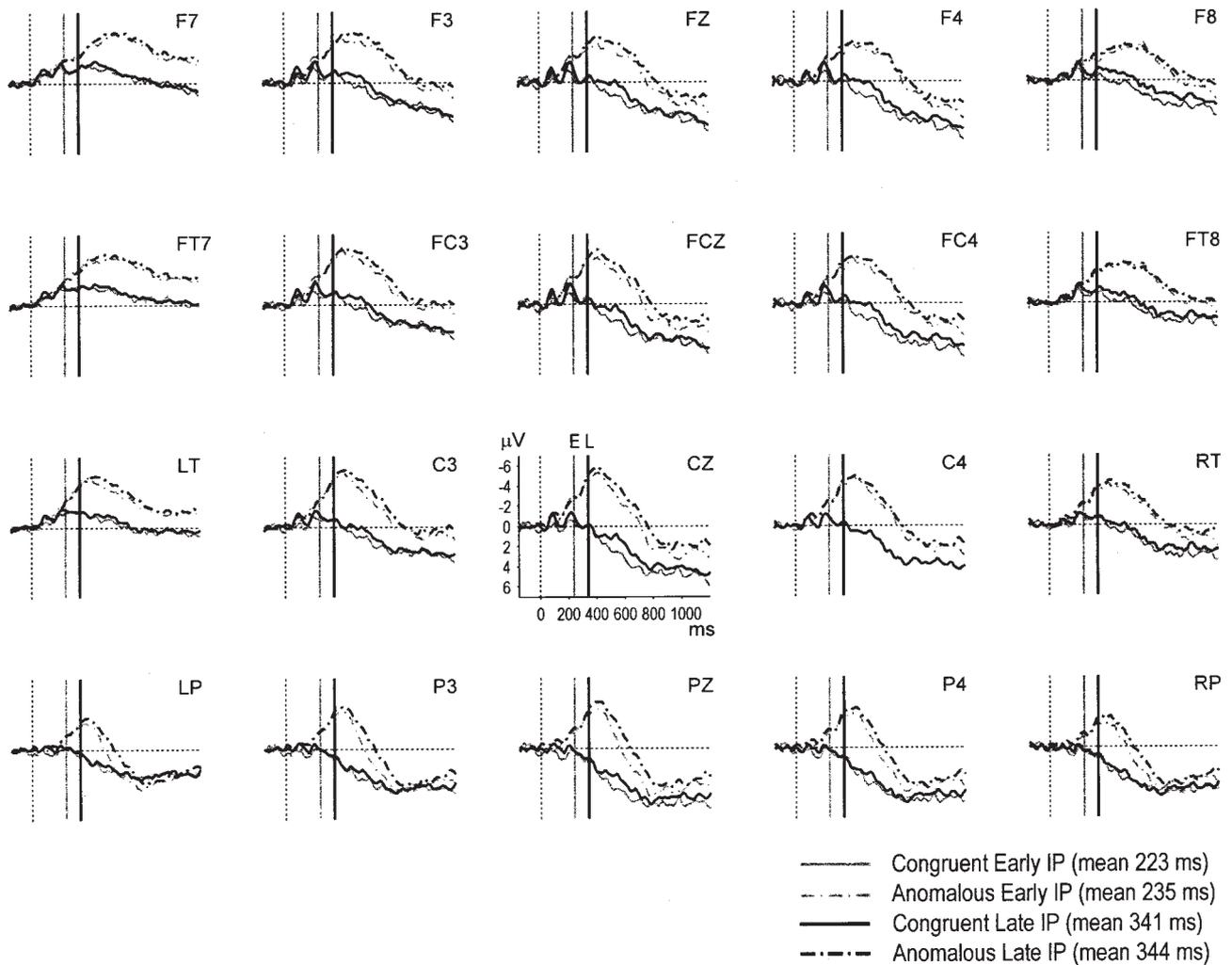
Figure 1. Connected speech. Grand average event-related potentials from 20 scalp sites to sentence-final words that were (a) congruent and had early isolation points (IPs; solid gray line), (b) semantically anomalous and had early IPs (gray alternating dash/dot line), (c) congruent and had late IPs (solid black line), and (d) semantically anomalous and had late IPs (black alternating dash/dot line), after baseline correction in the 150-ms prestimulus interval. Time 0 is the onset of sentence-final words. Lines E and L mark the mean IPs for the early and late comparisons. The time axis is in milliseconds. Note that negative polarity is plotted upward.

successfully singled out one candidate, integration processing was started up nonetheless. These results favor a cascading account of spoken-word processing in context. It seems that the semantic integration process does not wait until one appropriate candidate has been selected on the basis of a phonological analysis.

The results of the current study replicate earlier findings by Van Petten et al. (1999), which revealed that IP is not the critical time point for elicitation of the N400. The crucial difference between their study and ours is that our study is the first to use natural connected speech and manipulate IP to investigate the temporal relationship between word identification and semantic integration. Van Petten et al. concluded that onset of the N400 effect occurred not at IP but at the discrepancy point, the moment when the acoustic input first diverged from semantic expectations. This raises an interesting question about the cognitive mechanism that

drives language comprehension. Until recently, many researchers have shared the view that language processing is continuous and incremental but that each incoming word is processed with bottom–up priority, and a number of spoken word processing models, such as the Cohort model and Shortlist, have incorporated this view (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Norris, 1994). Our data could be explained in terms of bottom–up priority of spoken word processing. However, a number of recent ERP studies have investigated the possibility that expectancies are being generated on the basis of the preceding context in combination with the comprehender's knowledge about the world (De-Long, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). The results of these studies have revealed that in highly constraining contexts, such as the ones used in this study, words are not only

Table 3
ANOVA on Mean ERP Amplitude in the 300–500 ms Latency Range (N400)

Condition	Source	*df*s	*F*	*MSE*	*p*
Omnibus ANOVA (29 electrodes)					
Early IP: CC vs. IC	Con	1, 11	45.75	55.31	.000**
	Con × Elec	28, 308	9.52	0.68	.000**
Late IP: CC vs. IC	Con	1, 11	41.70	62.25	.000**
	Con × Elec	28, 308	10.66	0.78	.000**
Anterior vs. posterior (2 × 12 electrodes)					
Early IP: CC vs. IC	Con × Site	1, 11	8.96	4.08	.012*
Late IP: CC vs. IC	Con × Site	1, 11	8.59	7.03	.014*
Anterior left vs. anterior right (2 × 5 electrodes)					
Early IP: CC vs. IC	Con × Site	1, 11	1.63	1.39	.228
Late IP: CC vs. IC	Con × Site	1, 11	0.11	1.48	.745
Posterior left vs. posterior right (2 × 5 electrodes)					
Early IP: CC vs. IC	Con × Site	1, 11	0.91	2.07	.360
Late IP: CC vs. IC	Con × Site	1, 11	1.66	2.00	.223

Note. ANOVA = analysis of variance; ERP = event-related potential; IP = isolation point; CC = congruent condition; IC = incongruent condition; Con = congruity type; Elec = electrode.
* *p* < .05.   ** *p* < .001.

rapidly integrated into the higher order meaning representation of the preceding sentence or discourse context but that the constraining context is used to form probabilistic predictions of which specific words will be presented next. This idea of anticipation is not new. Several models of spoken word recognition, such as Morton's (1969) Logogen model and TRACE by McClelland and Elman (1986), already allowed for lexical preactivation of words on the basis of the context. However, until recently, compelling evidence for preactivation during online sentence processing has been lacking, whereas a number of behavioral studies have provided evidence for an initial bottom–up priority on the basis of the

acoustic input (for a review of the literature relevant to this issue, see Van Berkum et al., 2005).

On the basis of the results of the current study, it is impossible to determine whether the large N400 amplitude in the semantically anomalous condition reflects difficulty in integration or whether the N400 was elicited because prediction of specific words in these highly constraining sentence contexts was violated. In the case of a violation of anticipation of specific words, it is not surprising to find that the N400 effect sets in before IP. Analysis of the first phonemes reveals that they do not match with those of any of the words anticipated. However, in light of the majority of evidence

N400 effect Words with Early IP   **N400 effect Words with Late IP**

300 - 500 ms   300 - 500 ms

-6 µV   0 µV   -6 µV   0 µV
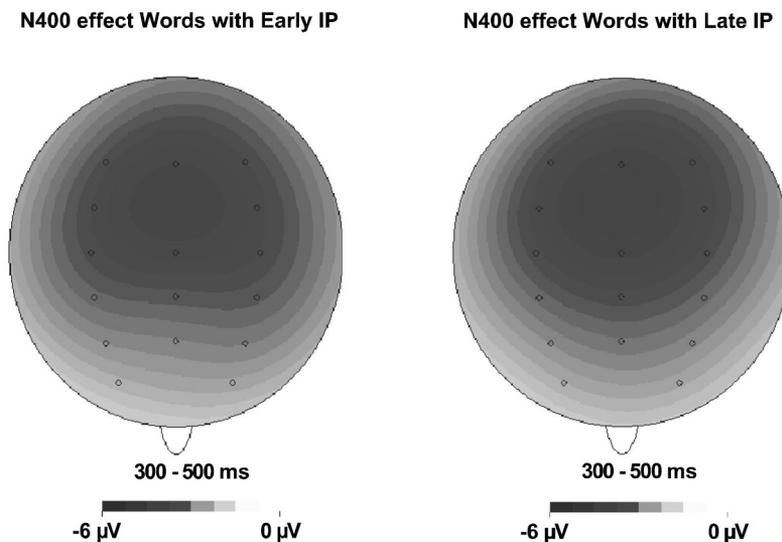
Figure 2. Connected speech. Isopotential voltage maps of the N400 effects for words with early isolation points (IPs) and words with late IPs. Maps are based on difference waveforms that resulted from subtracting the mean amplitude in the 300–500-ms latency window of the grand average event-related potential (ERP) for the congruent critical words from the mean amplitude of the grand average ERP for the semantically anomalous critical words.
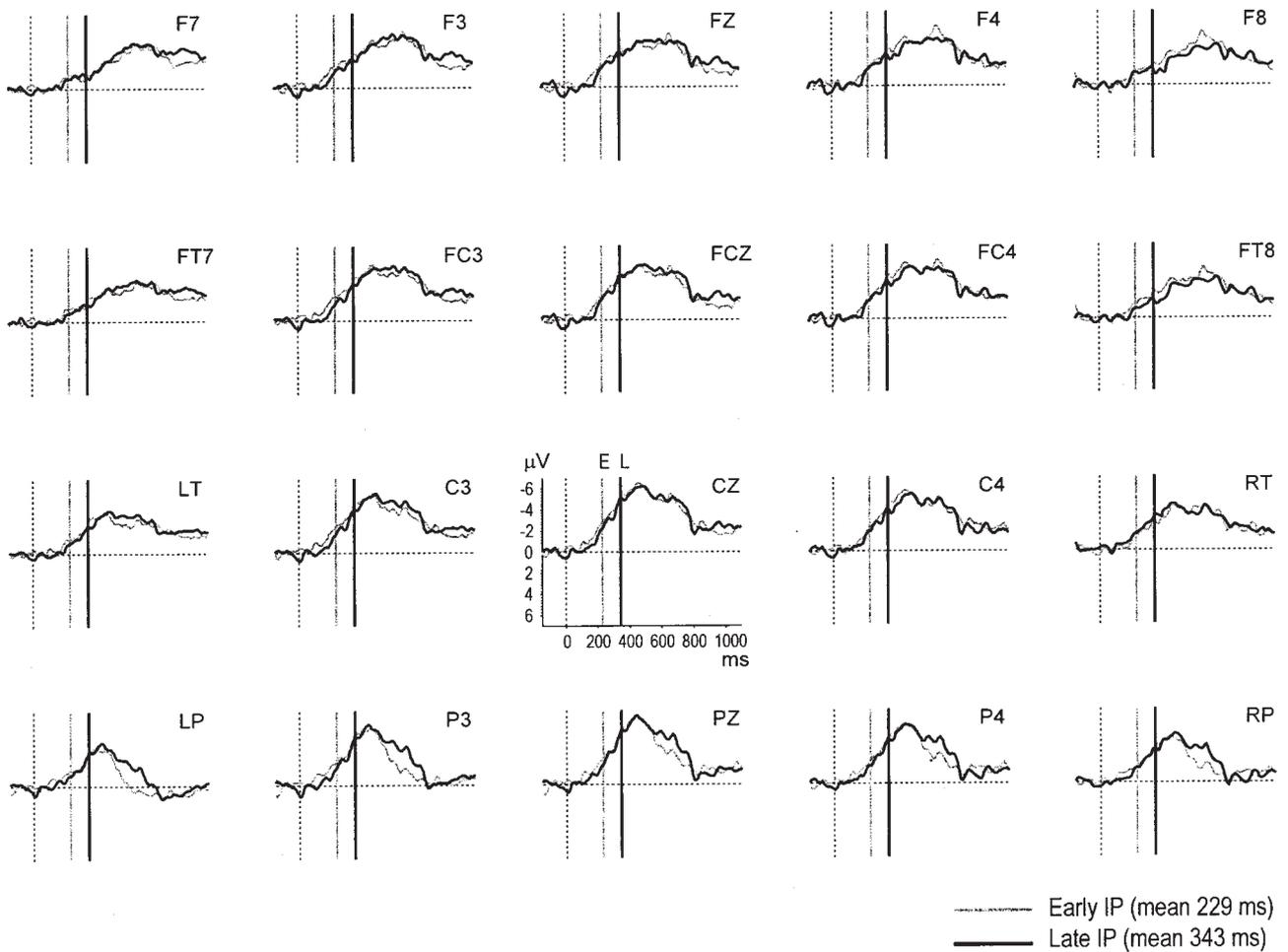
Figure 3. Connected speech. Difference waveforms from 20 scalp sites for semantically congruent versus semantically anomalous sentence-final words with early isolation points (IPs; solid gray line) and late IPs (solid black line), time-locked to IP. Lines E and L mark the mean IPs for the early and late comparisons. The time axis is in milliseconds. Note that negative polarity is plotted upward.

favoring bottom–up priority for acoustic processing of initial phonemes of the perceived word, we propose that our data suggest the following scenario. During sentence processing, there is a certain time frame in which lexical selection on the basis of a combination of the acoustic analysis of a word's first phonemes and context-based specifications can take place, and presumably does in the case of congruent words in highly to moderately constraining sentences. When, however, selection of one appropriate candidate is difficult, as would be the case for anomalous words, or even congruent words in low constraining contexts, integration as reflected by the N400 seems to be attempted for those candidates that match the acoustic input. Because speech is such a rapid phenomenon, one can imagine that the comprehension system cannot always wait for a word to be fully perceived. In these cases, integration would be attempted for a selected number of candidates that still match the acoustic input and, thus, would begin before word identification is complete. Common sense experience tells us that after unsuccessful integration of several candidates, the actual word eventually is perceived on the basis of

acoustic analysis alone, in which case identification of an anomalous word is usually closely followed by a "What did you say?" from the listener.

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

References

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

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., 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 neuropsychological responses to spoken sentences. *Brain and Language, 39,* 302–318.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbours of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501–518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1211–1238.

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

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

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

Kutas, M., & Hillyard, S. A. (1980, January 11). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

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

Maxwell, S. E., & Delaney, H. D. (1990). *Designing and analyzing data.* Pacific Grove, CA: Brooks/Cole.

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

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 621–638.

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

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

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1209–1228.

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

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research, 15,* 228–240.

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

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443–467.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17,* 701–718.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–985.

van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16,* 1068–1084.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394–417.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16,* 1272–1288.

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