

Event-Related Brain Potentials Elicited by Failure to Agree

LEE OSTERHOUT AND LINDA A. MOBLEY

University of Washington

Event-related potentials (ERPs) were recorded from 13 scalp electrodes while subjects read sentences, some of which contained violations of number or gender agreement. Subjects judged the acceptability of sentences in Experiments 1 and 2 and passively read sentences in Experiment 3. In Experiment 1, violations of subject–verb number, reflexive–antecedent number, and reflexive–antecedent gender agreement elicited a widely distributed positive-going wave (P600). Subject–verb agreement violations also elicited a left-hemisphere negativity. In Experiment 2, personal pronouns that mismatched in gender with the subject noun elicited a P600, but only when subjects judged such sentences to be unacceptable. Semantically anomalous words elicited an enhanced N400 component. In Experiment 3, subject–verb number disagreement elicited a P600 and semantic anomalies elicited an enhanced N400. ERPs to reflexive–antecedent agreement violations did not differ from those to controls. We evaluate the speculation that agreement between sentence constituents reflects syntactic constraints rather than semantic or discourse factors. © 1995 Academic Press, Inc.

In his introduction to transformational grammar, Radford (1988) examines the status of sentence (1):

- (1) The boy next door never loses her temper with anyone.

Although most people find such sentences unacceptable, Radford notes that the proper characterization of the oddity is open to interpretation. One could claim that the sentence is *syntactically anomalous*, i.e., that agreement in gender between a pronoun and its antecedent is stipulated as part of the syntactic constraints imposed by the grammar. Alternatively, one could claim that the sentence is *semantically anomalous*, i.e., that part of the meaning of the word *boy* denotes a male human, whereas part of the meaning of the pronoun *her* de-

notes a female human, leading to a contradiction in meaning if these two entities are taken to be coreferential. In the study reported here, we sought to investigate the following question: Is agreement in number and gender between sentence constituents determined by application of a set of formal syntactic rules, or by semantic properties or discourse functions of the entities referred to within the sentence? Such questions are traditionally the domain of linguistic theory. Viewed from the perspective of grammatical description, the problem is to determine (primarily through rational argument) which part of the grammar (e.g., syntactic or semantic) provides the most perspicuous account of agreement. The approach taken here is to obtain an empirical, evidential basis for answering this question, by observing the brain response to agreement violations encountered during language comprehension. This approach is motivated by recent evidence that, at least under certain conditions, certain types of syntactic and semantic anomaly elicit distinct patterns of brain response (cf. Osterhout, 1990; Osterhout & Holcomb, 1992, 1993; Osterhout, Holcomb, & Swinney, 1994).

Treatments of agreement within traditional grammars often distinguish between “controllers” (e.g., nouns in subject position) and

We thank Phil Holcomb, Jane Anderson, Janet Nicol, Dave Balota, and several anonymous reviewers for valuable comments on previous versions of this manuscript. Thanks also to Dawn Aiken, Rick McKinnon, Vicka Corey, Marty Blount, and Sherri Sipe for help in running subjects and preparing stimuli. This research was supported by National Institute of Deafness and Communication Disorders Grant DC01947 and by a grant from the University of Washington Graduate School Fund. Address correspondence and reprint requests to Lee Osterhout, Department of Psychology, University of Washington, Seattle, WA 98195.

"controllees" (e.g., tensed verbs and pronouns) (cf. Barlow & Ferguson, 1988). The standard view is that the grammatical number and gender of the controllee are mechanically "inherited" from the controller in a purely formal manner. This position is well represented by Jespersen (1933), who claimed that "the plural form of *sing* (in *birds sing*) . . . is only a meaningless grammatical contrivance showing the dependence of the verb on the subject." A typical formal rule accounting for subject-verb agreement in English is that "the verb that heads the predicate of a sentence must agree with the subject in number and person" (Baker, 1989, p. 38; see also Bach, 1974; Gleason, 1965; Palmer, 1978). This account makes use of the syntactic notions of grammatical role (i.e., the notion of subject) and hierarchical structure (the subject is identified as the noun phrase dominated by the sentence symbol). Current grammars also place number and gender agreement within a theory of syntax, representing agreement as a feature bundle residing under an inflectional node within the phrase structure of the sentence (cf. Aissen, 1988; Chomsky, 1981, 1986; Doron, 1988; Radford, 1988; Sells, 1985). Number, gender, and other agreement properties are then mechanically "copied" from the controller to the controllee.

This standard view stands in striking contrast to proposals that number, gender, and related agreement phenomena are largely governed by *semantic* and *discourse* factors (Barlow, 1988; Chafe, 1970; Croft, 1988; Fassi Fehri, 1988; Lapointe, 1988; Pollard & Sag, 1988; Reid, 1991; Wierzbicka, 1985). For example, Reid (1991) questions the claim that agreement between sentence constituents is determined strictly by formal rule. If such were the case, then one would expect the grammatical form of the controllee to predictably follow from the grammatical form of the controller. However, Reid (1991) claims that the putative agreement rules are broken with regularity, both in print and in speech. Consider (2a) and (2b):

(2a) My *family have* been prominent, well-to-do people in this Middle Western city for three generations.

(2b) Her *family is* one aunt about a thousand years old. (F. Scott Fitzgerald, *The Great Gatsby*)

Reid accounts for such inconsistencies by positing a "Focus Number System," in which verbal suffixes indicating number are chosen to correspond with the entity in discourse focus, rather than being determined entirely by the form of the subject noun. In (2a), the focus is on a group of well-to-do people, whereas in (2b) the focus is on a single person. Correspondingly, the plural form of the verb is employed in (2a), while the singular form is employed in (2b).

A related question is whether there is one type of agreement relation or many types (cf. Ferguson & Barlow, 1988). In particular, one can ask whether subject-verb agreement and agreement between anaphors and their antecedents should be treated as distinct phenomena. Some linguists (e.g., Givon, 1976, 1979) have argued that subject-verb agreement is diachronically and functionally related to anaphora. Both are presumed to play a discourse function by indicating salient aspects of the discourse. However, a striking observation of a dissociation in the agreement preferences for verbs and anaphors indicates that different types of information might control these two forms of agreement. Speakers of American English typically prefer to use a singular verb after collective nouns (e.g., "The committee *is* meeting.,"; Huddleston, 1984). At the same time, such speakers judge plurals as more natural anaphors for collectives (e.g., "The committee *is* meeting. *They* will discuss the proposal.") and they read sentences with plural anaphors for collectives more rapidly than the same sentences with singular anaphors (Gernsbacher, 1991). One explanation for this discrepancy is that the "grammatical" singular property of collective nouns governs subject-verb agreement, whereas the "notional" or "semantic" sense of these nouns governs anaphoric agreement (cf. Bock & Eberhard, 1993).

Thus, the question of whether agreement in number and gender is best thought of as being part of the form or function of language re-

mains a point of debate. We addressed this question by recording event-related brain potentials (ERPs) elicited during sentence processing. ERPs, which are non-invasive, continuous, and on-line measures of the electrical activity of the brain, have been shown to be sensitive to a variety of sensory and cognitive processes (for review, Hillyard & Picton, 1987). Of particular relevance to the current study is recent evidence that ERPs are differentially sensitive to syntactic and semantic aspects of comprehension (Hagoort et al., 1993; Osterhout, 1990, in press; Osterhout & Holcomb, 1992, 1993; Osterhout, Holcomb, & Swinney, 1994). In a pioneering set of experiments, Kutas and Hillyard reported that one type of semantic anomaly (contextually inappropriate words, e.g., "I take my coffee with cream and *dog*") elicits a greatly enhanced negative component with a peak around 400 ms poststimulus (the N400 component; Kutas & Hillyard, 1980a, 1980b, 1980c; for review, Kutas & Van Petten, 1988). Typically, this "N400 effect" is largest posteriorly and over the right hemisphere. The precise cognitive events underlying the N400 are not known. One suggestion is that N400 amplitude is inversely related to the amount of lexical priming impinging on the representation of the target word from preceding context (Holcomb & Neville, 1990). Another proposal is that N400 amplitude reflects the buildup of semantic constraints imposed by preceding context (Van Petten & Kutas, 1990). In any case, there is a consensus that N400 amplitude is a function of the "semantic fit" between the target word and preceding context (cf. Kutas, 1993).

More recently, other researchers have examined the ERP response to violations of syntactic constraints. Critically, such violations do not elicit the "classical" N400. At least under certain experimental conditions, violations of phrase structure (Hagoort, Brown, & Groothusen, 1993; Neville, Nicol, Barrs, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992, 1993), verb subcategorization (Osterhout & Holcomb, 1992; Osterhout et al., 1994), and subadjacency (McKinnon & Osterhout, 1995; Neville et al., 1991) have been

found to elicit a large positive-going wave with an onset around 500 ms. This effect has been labelled the "P600" by Osterhout and Holcomb (1992) and the "syntactic positive shift" by Hagoort et al. (1993). In some reports this positivity is preceded by a left-hemisphere negativity between roughly 300 and 500 ms (Neville et al., 1991; Osterhout & Holcomb, 1992). Although the precise cognitive events underlying these language-sensitive effects remain unknown, it is clear that the P600 (the response to a variety of sentence-embedded syntactic anomalies) and the N400 (the response to semantically anomalous words) are neurally (and, by inference, cognitively) distinct. One goal of the current study, then, was to determine whether the brain response to agreement violations most closely resembles the P600 or N400 response.

Previous experiments examining the ERP response to agreement violations have produced conflicting results. Kutas and Hillyard (1983) presented sentences containing errors involving noun number, verb number, and verb tense. Few reliable differences were found in ERPs elicited by violating and control words, although the errors were associated with increased negativity between 200 and 500 ms poststimulus at some (mostly anterior) sites. More recently, Hagoort et al. (1993) observed the response to agreement violations in Dutch. The critical words, always verbs, disagreed with the subject noun in number. Hagoort et al. observed a positive-going shift with an onset around 500 ms, similar to the P600 effect reported by Osterhout and Holcomb (1992). The reason for this discrepancy is not clear, although one possibility is that it is related to differences between English and Dutch. The Dutch agreement system is considerably more extensive than the English system (involving, e.g., agreement markers on articles and adjectives). Bates, MacWhinney, and colleagues have suggested that comprehenders' sensitivity to a particular cue may reflect the relative importance or salience of that cue (Bates & MacWhinney, 1987; MacWhinney, Bates, & Kliegl, 1984). Hence, one speculation is that the large brain response to agreement violations

shown by Dutch comprehenders, and the apparent absence of such a response in English comprehenders, reflects the relative "salience" or importance of agreement cues in the two languages. Another explanation notes that Kutas and Hillyard (1983) plotted only the first 600 ms or so of activity following onset of the anomalous word. Given that the P600 effect reported in other studies typically has an onset around 500 ms and a duration of at least several hundred ms, it is conceivable that Kutas and Hillyard did not plot a sufficiently long portion of the waveform to observe a P600-like effect even if it was present in their data.

To summarize, the current study was designed to investigate three questions. First, is there a reliable and measurable brain response to violations of number and gender agreement? Second, does this response more closely approximate the response to semantically anomalous words (N400) or violations of phrase structure, subcategorization, and constituent movement constraints (P600)? And third, is the response to violations of subject-verb agreement similar or dissimilar to violations of anaphoric agreement?

EXPERIMENT 1

Method

Subjects

Sixteen undergraduates from the University of Washington (eight males and eight females) participated for course credit. Ages ranged from 18 to 32 (mean = 21.5) years. In this and all subsequent experiments, all subjects were right-handed (Edinburgh inventory) native English speakers with normal or corrected-to-normal vision.

Materials

Three sets of 60 experimental sentence pairs were constructed, as exemplified by sentences (1)–(3) in Table 1. Sentences in the *subject-verb agreement condition* began with a plural noun phrase in subject position. In one version of each pair, the verb appeared in its plural form (subject-verb number agreement). In

TABLE 1
EXAMPLES OF EXPERIMENTAL SENTENCES
PRESENTED IN EXPERIMENT 1

1. The elected officials *hope/hopes* to succeed.
(subject-verb number agreement condition)
2. The hungry guests helped *themselves/himself*
to the food.
(reflexive-antecedent number agreement condition)
3. The successful woman congratulated *herself/himself*
on the promotion
(reflexive-antecedent gender agreement condition)

the second version of each pair, the verb appeared in its singular form (subject-verb number disagreement).¹ Sentences in the *reflective-antecedent number agreement condition* began with a definite noun phrase in subject position. Half of these sentences contained a plural subject noun and half contained a singular subject noun. These sentences also contained a reflexive (*himself*, *herself*, *themselves*) acting either as object of the verb or as object of a preposition. In all cases, the antecedent to the reflexive was the noun phrase in subject position. In half of the sentences (15 with singular subject nouns and 15 with plural subject nouns), the reflexive matched its antecedent in number, whereas in the remaining sentences the reflexive and its antecedent were discordant in number. Finally, sentences in the *reflexive-antecedent gender agreement condition* contained a singular definite noun phrase in subject position. These nouns were occupations, titles, or states that unambiguously indicated feminine or masculine (as determined by the experimenters' judgments; e.g., *waitress*, *queen*, *bachelor*). Half of the subject nouns were feminine and half were masculine. These sentences also contained a reflexive pronoun in singular form (*himself* or *herself*) acting as object of the verb or a preposition. Again, the antecedent of the reflexive was always the noun

¹ We chose not to examine the response to correct and incorrect verb number agreement for singular-subject nouns, because the nature of the anomaly is often ambiguous when the proper marker is missing. For example, the anomaly in the sentence "The actress hope . . ." might result from the absence of a past tense marker or from the absence of a singular number marker.

in subject position. In half of the sentences (15 with feminine subject nouns and 15 with masculine subject nouns) the reflexive matched its antecedent in gender; in the remaining half it did not. The complete set of experimental sentences is provided in Appendix 1.

In addition to the experimental sentences, a set of 30 filler sentences was constructed. Ten of these sentences were similar in form to the experimental sentences in the subject-noun agreement condition, except that they contained a singular noun in subject position. Five of these contained a singular verb, and 5 contained a plural verb. The 20 remaining fillers (10 grammatical, 10 containing phrase structure violations) began with a singular definite noun phrase and did not contain a reflexive pronoun. Hence, across all of the materials, 105 sentences were well formed and 105 were ill formed. The above materials were counterbalanced across two stimulus lists, such that each list contained one version of each sentence pair, and 30 exemplars of each sentence type.

Procedure

Each trial consisted of the following sequence of events: A fixation cross appeared for 500 ms, after which a sentence was presented in a word-by-word manner, with each word appearing approximately in the center of the screen for 350 ms. A blank screen inter-stimulus interval of 300 ms separated words.² Sentence-final words appeared with a period. A 1400-ms blank screen interval followed each sentence, after which a prompt appeared on the screen asking subjects to respond by pressing one of two buttons. The prompt remained on the screen until the subject responded, after which the next trial was initiated.

Subjects were tested in one session that lasted from one to two hours, during which they were seated in a comfortable chair situated in an isolated room. Subjects were instructed to carefully read each sentence, and also to judge the "acceptability" of each sentence. Unac-

ceptable sentences were defined as those that contained *any type* of anomaly (e.g., oversized words, ungrammaticalities, semantic anomalies, etc.). It was stressed to the subjects that they were *not* to make grammaticality judgments. Subjects were asked if they understood these criteria, and examples (not from the stimulus list) were provided as needed. The buttons used to indicate "acceptable" and "unacceptable" (left and right hand) were counterbalanced across subjects.

Data Acquisition and Analysis

Electroencephalographic (EEG) activity was recorded from 13 scalp locations using tin electrodes attached to an elastic cap (Electrocap International). Electrode placement included International 10–20 system locations (Jasper, 1958) at homologous positions over the left and right occipital (O1, O2) and frontal (F7, F8) regions, and from frontal (Fz), central (Cz), and parietal (Pz) midline sites. In addition, several non-standard sites over posited language centers were used, including Wernicke's area (WL, WR; 30% of the interaural distance lateral to a point 13% of the nasion-inion distance posterior to Cz), temporal (TL, TR; 33% of the interaural distance lateral to Cz), and anterior temporal (ATL, one-half the distance between F7 and T3; ATR, one-half the distance between F8 and T4). Vertical eye movements and blinks were monitored by means of an electrode positioned beneath the left eye. Horizontal eye movements were monitored by an electrode positioned to the right of the right eye. The above 15 channels of EEG were referenced to the left mastoid and were amplified with a bandpass of 0.01 to 100 Hz (3db cutoff) by a Grass Model 12 amplifier system. Activity over the right mastoid was actively recorded on a sixteenth channel in order to determine if there were any effects of the experimental variables on the mastoid recordings. No such effects were observed in any of the analyses described below. Electrode impedance was kept below 5 kohm at all scalp and mastoid sites and below 20 kohm for the eye electrodes.

Continuous analog-to-digital conversion of

² We used a 650-ms SOA between words so that we could examine an extended period of ERP activity to each word, uncontaminated by the ERP to the next word.

the EEG and stimulus trigger codes was performed on-line by a Data Translation 2801-A board and an AT-compatible computer, at a sampling rate of 200 Hz. Trials characterized by excessive eye movement or amplifier blocking were rejected. ERPs were quantified by computer as the mean voltage within a latency range following presentation of the word of interest, relative to 100 ms of activity prior to onset of the word of interest. We performed analyses on mean voltage within four latency windows: 50–150, 150–300, 300–500, and 500–800 ms. These windows were chosen because they roughly correspond to the temporal windows associated with the N1, P2, N400, and P600 components often elicited by linguistic stimuli. ERPs were quantified using a baseline composed of the 100 ms of activity preceding onset of the critical word. Data acquired at midline and lateral sites were treated separately to allow quantitative analysis of hemispheric differences. Two-way ANOVAs with repeated measures on sentence type (agreement violating vs non-violating) and electrode position (frontal, central, and posterior) were performed on data acquired at midline sites. Three-way ANOVAs with repeated measures on sentence type, hemisphere, and five electrode positions were performed on data acquired over lateral sites.

Finally, the grand averages reported here were not response-contingent. All artifact-free trials for a given condition were entered into the grand average, regardless of the subject's response on that trial. We adopted this strategy because we do not know what the relationship is between the subjects' delayed, end-of-sentence judgment and the immediate processing response to the critical words.

Results

Acceptability Judgments

The mean percentages of trials for each agreement condition on which subjects judged non-violating and agreement-violating sentences to be acceptable were as follows: subject–verb number, 86 and 12%; reflexive–antecedent number, 82 and 11%; and reflexive–antecedent gender, 87 and 10%. In each condition, the difference in proportion of agreement-violating and non-violating sentences judged to be acceptable was statistically reliable ($p < .0001$ in each comparison).

ive–antecedent gender, 87 and 10%. In each condition, the difference in proportion of agreement-violating and non-violating sentences judged to be acceptable was statistically reliable ($p < .0001$ in each comparison).

Event-Related Potentials

ERPs to verbs and reflexives. Thirteen percent of the trials involving verbs and reflexives were rejected for artifact prior to averaging. These trials were randomly distributed across conditions. Responses to the verbs and reflexives (averaged across subjects and items) in the subject–verb number agreement, reflexive–antecedent number agreement, and reflexive–antecedent gender agreement conditions are shown in Figs. 1, 2, and 3, respectively. In these and all subsequent figures, the general shapes of the waveforms were consistent with those reported in other experiments involving language stimuli (Kutas & Hillyard, 1980c; Neville, Kutas, Chesney, & Schmidt, 1985; Osterhout & Holcomb, 1992). A clear negative-positive complex was visible in the first 300 ms after stimulus presentation. The negative component ("N1") peaked at about 120 ms at most sites, and was larger at midline and anterior than at posterior regions. The positive component ("P2") peaked at about 220 ms and was largest over midline sites. A large-amplitude negative deflection with a centro-posterior distribution and a peak amplitude around 400 ms (N400) was also clearly visible in the ERPs to most words. Sentence-final words were followed by a large-amplitude positivity, which is often observed following sentence-final words (e.g., Friedman, Simson, Ritter, & Rapin, 1975).

The most notable difference between the ERPs to subject–verb number agreement violations and controls was the presence of a large, widely distributed positive-going wave with an onset around 500 ms (Fig. 1). Smaller differences were also present in the waveforms preceding the onset of this positivity. In particular, P2 amplitude was slightly larger in the agreement-violating condition than in the non-violating condition, particularly at posterior and right-hemisphere sites. Additionally, ERPs

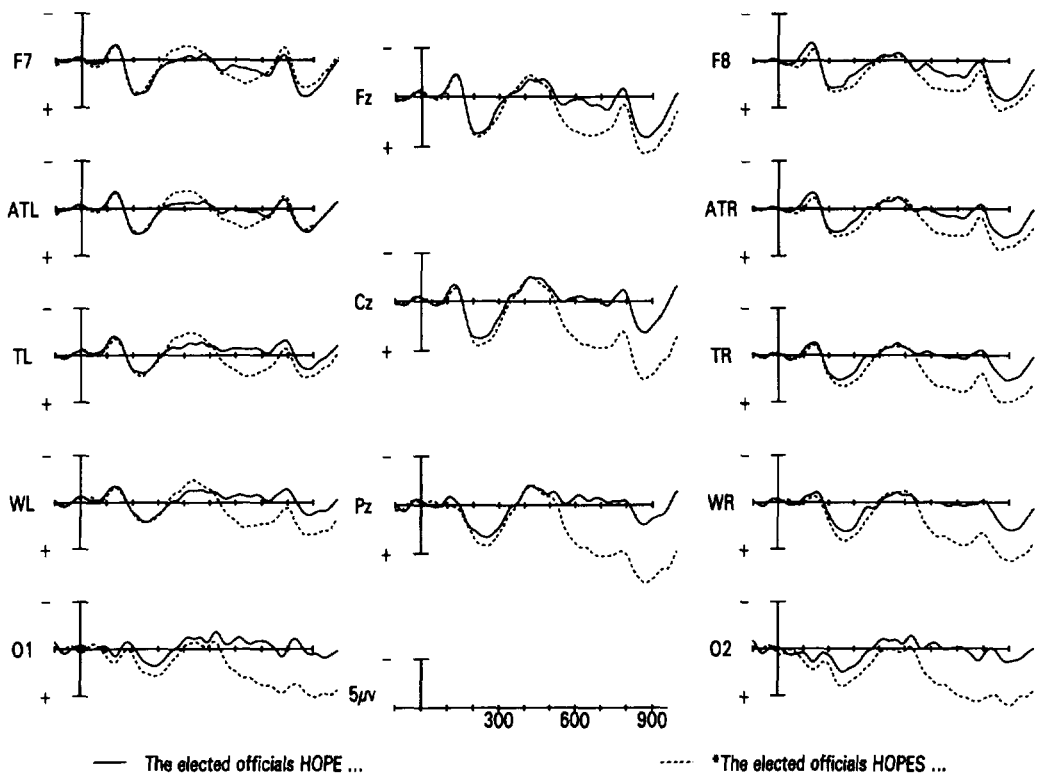


FIG. 1. Grand average ERPs recorded over three midline and 10 lateral sites to subject-verb number violations and controls. Onset of the critical words in non-violating (solid line) and agreement-violating (dashed line) conditions is indicated by the vertical bar. Each hash mark represents 100 ms. Positive voltage is plotted down.

to agreement-violating verbs were more negative-going over anterior and temporal left hemisphere sites between roughly 300 and 500 ms. Statistical analyses confirmed these observations. In these and all subsequent analyses, the Geisser-Greenhouse (1959) correction was applied when evaluating effects with more than one degree of freedom in order to protect against excessive Type I error due to violations of the assumption of equal variances between conditions of within-subject factors. Within the 150–300 ms window, no reliable differences were observed between conditions over midline sites. However, reliable interactions at lateral sites between sentence type and hemisphere ($F(1,15) = 7.24, p < .05, MS_e = 0.75$) and between sentence type and electrode site ($F(4,60) = 3.22, p = .06, MS_e = 0.83$) reflected the fact that agreement violations elicited a large P2 amplitude at some sites than did

controls. Within the 300–500 ms window, ERPs to agreement-violating verbs were more negative-going, but only at anterior and temporal sites in the left hemisphere (sentence type \times electrode site, $F(4,60) = 3.20, p = .07, MS_e = 1.47$; sentence type \times hemisphere: $F(1,15) = 9.88, p < .01, MS_e = 2.32$). The large positive-going activity elicited by the agreement-violating words was highly reliable in the 500–800 ms window (midline, $F(1,15) = 15.75, p < .01, MS_e = 19$; lateral, $F(1,15) = 12.40, p < .01, MS_e = 26.69$), and differences between conditions were largest at posterior sites (sentence type \times electrode site: midline, $F(2,30) = 6.21, p < .01, MS_e = 1.48$; lateral, $F(4,60) = 7.55, p < .01, MS_e = 2.61$).

ERPs to reflexives are shown in Figs. 2 (number condition) and 3 (gender condition). The only reliable effect within the first three windows (50–150, 150–300, and 300–500 ms)

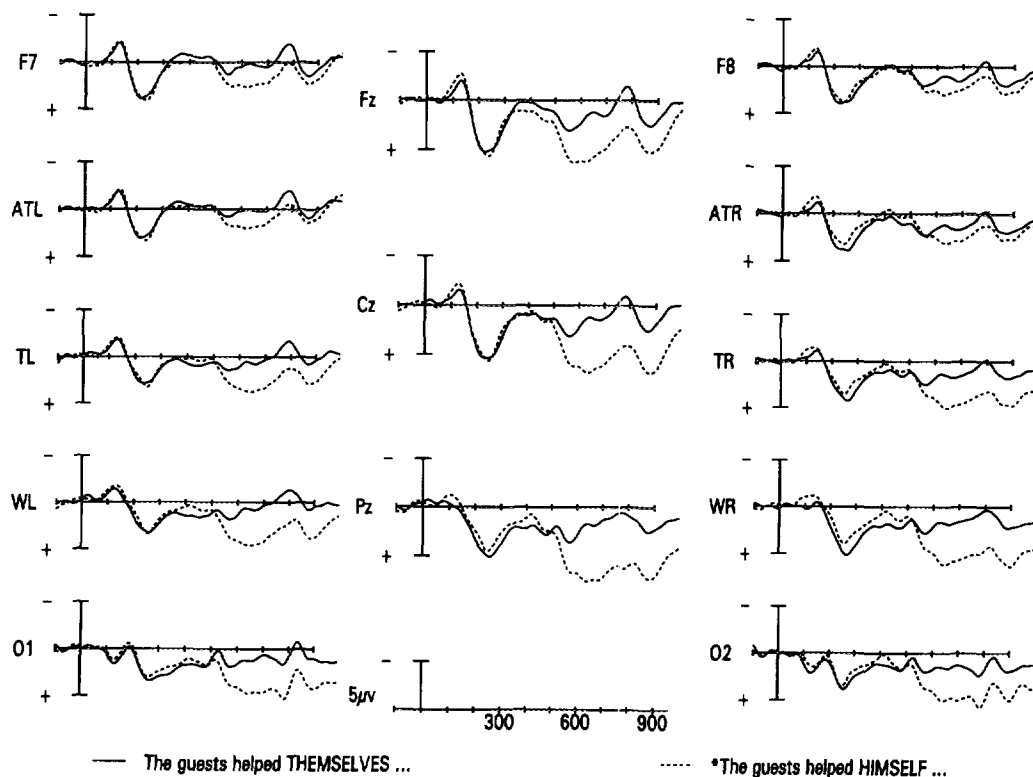


FIG. 2. Grand average ERPs to reflexive-antecedent number violations (dashed line) and controls (solid line).

in either reflexive condition was an interaction between sentence type and electrode site at midline sites from 300 to 500 ms in the number agreement condition, $F(2,30) = 10.34$, $p < .001$, $MS_e = 0.40$. This interaction reflected the fact that ERPs to agreement-violating reflexives were slightly more positive-going than controls at Fz and slightly more negative-going at Pz. However, subsequent pairwise comparisons of mean voltage at each site found no reliable differences at any electrode site, $p > .15$ in all comparisons. Differences between conditions within the 500–800 ms window (encompassing the large positive-going wave to agreement violations) were highly reliable, both in the number (midline, $F(1,15) = 43.83$, $p < .0001$, $MS_e = 9.05$; lateral, $F(1,15) = 47.13$, $p < .001$, $MS_e = 7.97$) and gender (midline, $F(1,15) = 20.24$, $p < .001$, $MS_e = 12.96$; lateral, $F(1,15) = 10.36$, $p < .01$, $MS_e = 20.90$) agreement conditions. Furthermore, dif-

ferences at lateral sites tended to be larger over posterior regions (sentence type \times electrode site: number, $F(4,60) = 5.27$, $p < .05$, $MS_e = 2.16$; gender, $F(4,60) = 6.19$, $p < .01$, $MS_e = 1.79$).

ERPs to sentence-final words. Osterhout and Holcomb (1992, 1993) previously reported that sentence-final words in sentences typically judged to be unacceptable elicit an enhanced negativity similar to the N400 effect. ERPs to sentence-final words in the subject-verb number, reflexive-antecedent number, and reflexive-antecedent gender conditions are plotted in Figs. 4A, 4B, and 4C, respectively. (Twenty-four percent of the trials involving sentence-final words were rejected for artifact.) ERPs to the sentence-final words in the ill-formed sentences were more negative-going than those to well-formed sentences, most notably between about 300 and 500 ms. Statistical analyses revealed no reliable differences between condi-

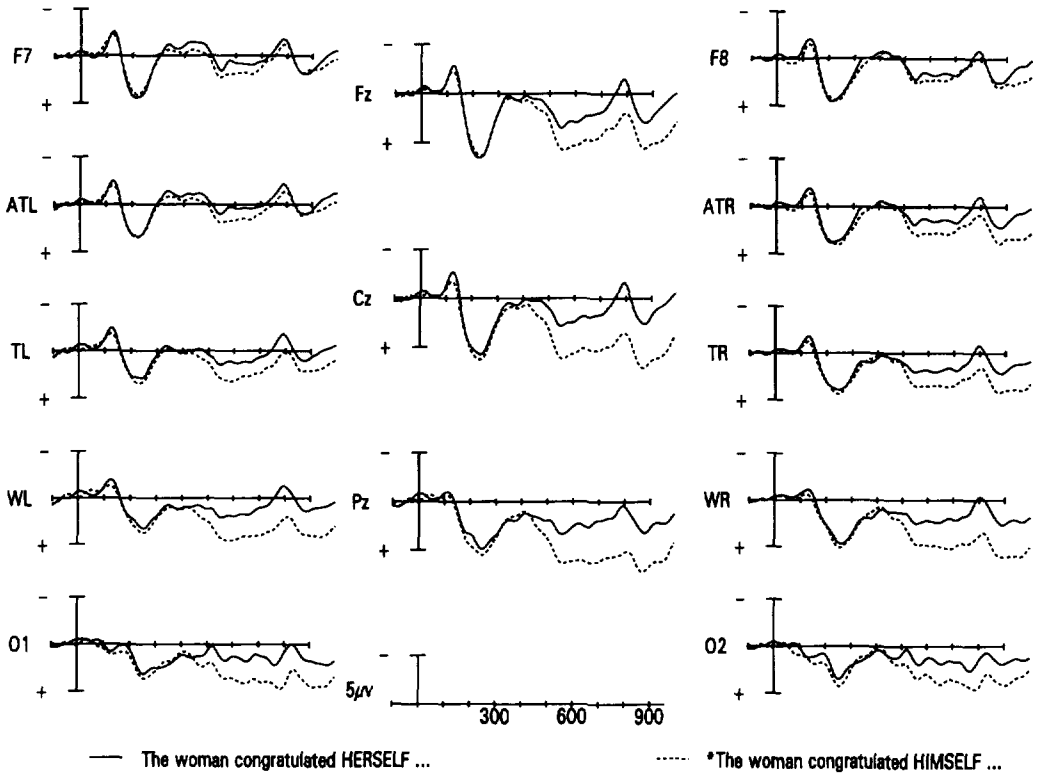


FIG. 3. Grand average ERPs to reflexive-antecedent gender violations (dashed line) and controls (solid line).

tions within the 50–150 and 150–300 ms windows in any of the comparisons. Between 300 and 500 ms, ERPs to agreement-violating words were more negative-going than controls

for all three agreement conditions, both at mid-line (subject–verb number: $F(1,15) = 13.80$, $p < .01$, $MS_e = 17.59$; reflexive–antecedent number: $F(1,15) = 10.23$, $p < .001$, $MS_e =$

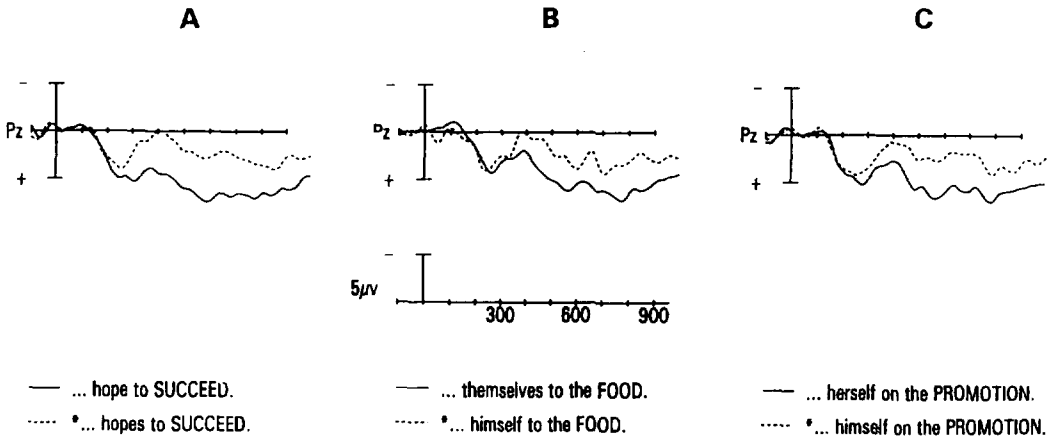


FIG. 4. Grand average ERPs to sentence-final words (recorded over site Pz) in sentences containing agreement violations (dashed line) and non-violating control sentences (solid line). (A) subject–verb number condition. (B) reflexive–antecedent number condition. (C) reflexive–antecedent gender condition.

8.90; reflexive-antecedent gender: $F(1,15) = 21.71, p < .001, MS_e = 10.24$) and at lateral sites (subject-verb number: $F(1,15) = 39.41, p < .0001, MS_e = 10.27$; reflexive-antecedent number: $F(1,15) = 9.99, p < .01, MS_e = 11.96$; reflexive-antecedent gender: $F(1,15) = 8.46, p < .01, MS_e = 30.04$). These effects were largest posteriorly in the subject-verb number condition (sentence type \times electrode site: midline, $F(2,30) = 3.97, p < .05, MS_e = 0.95$; lateral: $F(4,60) = 8.77, p < .01, MS_e = 1.03$) but were more evenly distributed across the scalp in the other conditions.

These effects extended into the 500–800 ms window. ERPs to agreement-violating words were more negative-going (or less positive-going) than those to controls in all three conditions at midline (subject-verb number: $F(1,15) = 9.37, p < .01, MS_e = 22.74$; reflexive-antecedent number: $F(1,15) = 10.56, p < .01, MS_e = 12.68$; reflexive-antecedent gender: $F(1,15) = 21.75, p < .001, MS_e = 13.60$) and at lateral sites (subject-verb number: $F(1,15) = 8.77, p < .01, MS_e = 22.77$; reflexive-antecedent number: $F(1,15) = 6.84, p < .05, MS_e = 16.24$; reflexive-antecedent gender: $F(1,15) = 7.95, p < .01, MS_e = 48.08$). Again, differences between conditions were largest posteriorly in the subject-verb number condition (sentence type \times electrode site; midline: $F(2,30) = 6.54, p < .05, MS_e = 2.07$; lateral: $F(4,60) = 7.64, p < .01, MS_e = 2.41$).

Discussion

All three types of agreement violations elicited a wide-spread, large-amplitude positive-going wave with an onset around 500 ms. This effect was highly similar to the P600 previously observed in response to a variety of syntactic anomalies (Hagoort et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1992, 1993; Osterhout et al., 1994). In addition, subject-verb number agreement violations elicited an enhanced left-hemisphere anterior and temporal negativity between roughly 300 and 500 ms and an increase in the P2 component, effects that have also been previously reported (Neville et al., 1991; Osterhout & Holcomb, 1992). These data could be taken to indicate that syntactic anomalies and agreement viola-

tions engender a similar processing response, one that is distinct from that engendered by semantically inappropriate words. Correspondingly, one might conclude that agreement violations are perceived to be syntactic in nature.

However, there are at least three objections to such a conclusion. The first objection concerns the possibility that the response to the anomaly has been confounded with the effects of the sentence-acceptability judgment task. The amplitude of a well-known positive-going ERP component, P300, varies as a function of the unexpectedness and task-relevance of the eliciting stimulus (Donchin, 1981; Donchin & Coles, 1988). One could reasonably argue that the agreement-violating words (which clearly indicated the proper response in the judgment task) were less expected and more task-relevant than the non-anomalous control words (which did not indicate the proper response), and therefore elicited a larger-amplitude P300.

A second (related) objection concerns the existence of differences between conditions that preceded the onset of the large positivity. In particular, ERPs to agreement violations were more negative-going than controls at some sites between 300 and 500 ms, the window normally associated with the N400 component. Although these effects were generally statistically unreliable (excepting the left hemisphere negativity to subject-verb number agreement violations) and did not uniformly have the centro-posterior distribution characteristic of the N400 effect, it is conceivable that these effects represent the remnants of more robust N400 effects that were partially obscured due to the effects of component overlap. That is, the large P300/P600 might have obscured an increase in N400 amplitude. Because the positivity was maximal over posterior regions, the effect of the overlap might have been to more dramatically reduce the size of the N400 effect over posterior than over anterior sites, leading to an unusual scalp distribution for the N400 effect.

One means for evaluating these possibilities is to observe the response to semantically anomalous words under the conditions employed in Experiment 1. If the positive-going wave reflects the greater unexpectedness and

task relevance of the anomalous words, relative to controls, then *any type* of unexpected, task-relevant anomaly should elicit a similar positivity. Furthermore, if the positivity obscured large-amplitude N400 effects in the response to agreement violations, then a similar obscuring should occur in the case of semantically anomalous words, which are known to elicit large N400s. In Experiment 2, we tested these possibilities by presenting semantically anomalous words under experimental conditions identical to those used during Experiment 1.

A third objection stems from the existence of an alternative interpretation of these data, at least for sentences containing reflexive pronouns. Within the current Government and Binding (GB) theory of grammar (Chomsky, 1981, 1986), the coindexation of reflexives and antecedents is governed by a set of formal rules specified within a theory of "binding." One such constraint is that a reflexive and its antecedent *must* be clausemates. (Within the GB framework, this is referred to as Principle A of binding theory.) Recent work has been taken to indicate that the comprehension system quickly discards as candidate antecedents any nouns that mismatch in number or gender with the reflexive (Corbett & Chang, 1983; Nicol, 1988). Since the subject noun was the only noun clausemate of the reflexive, such a strategy would result in a null set of candidate antecedents. This would force the comprehension system to search outside the current clause for an antecedent, violating Principle A. If the P600 elicited by reflexives in the agreement-violating conditions was in fact elicited in response to a perceived violation of Principle A of binding theory rather than in response to the mismatch in gender or number, then the presence of the P600 to such anomalies would not necessarily indicate that the agreement constraints themselves are encoded within the formal system. We examined this possibility in Experiment 2.

EXPERIMENT 2

In Experiment 2, we examined the response to agreement mismatches involving personal pronouns. In contrasts to the constraints associated with reflexive pronouns, binding theory

specifies only that the pronoun and its antecedent *cannot* be clausemates (Principle B). Hence, a pronoun may, but need not, be bound to an explicitly mentioned antecedent that is located outside of the clause containing the pronoun. We presented sentences containing both a noun indicating an unambiguously male or female occupation or title in subject position of the matrix clause, and a pronoun acting as subject of an embedded clause (Table 2). The pronouns in such sentences would be anomalous (i.e., a violation of gender agreement) only if the pronoun and the subject noun are perceived to be coreferential. The resulting anomaly could not be construed as a violation of binding theory, since no binding constraints would be violated regardless of whether or not the pronoun and subject noun are taken to be coreferential. Such an interpretation would, however, violate the putative formal rule that pronouns and antecedents must agree in gender.

We also presented sentences containing a semantically anomalous word. If the P600 is simply a P300-like response to the unexpectedness and task relevance of linguistic anomalies, then semantic anomalies should elicit a similar response. Furthermore, if this positivity acted to obscure an increase in N400 amplitude to syntactic anomalies due to component overlap, then the increase in N400 known to occur in response to semantically anomalous words should be similarly obscured in Experiment 2.

Method

Subjects

Twelve undergraduates (eight females and four males) participated for course credit. Ages ranged from 18 to 26 years (mean = 21 years).

Materials

Sixty sentence pairs were constructed, each

TABLE 2
EXAMPLES OF EXPERIMENTAL SENTENCES
PRESENTED IN EXPERIMENT 2

- | |
|---|
| 1. The aunt heard that <i>she/he</i> had won the lottery.
gender match/mismatch pronoun sentences |
| 2. The boat sailed down the river and <i>sank/barked</i>
during the storm.
non-anomalous/semantically anomalous sentences |

of which contained a personal pronoun in subject position within an embedded clause. The subject of the matrix clause was always inherently male or female (again as determined by experimenter judgment), with equal numbers of "male" and "female" nouns across all sentences. In one version of each sentence pair (the "gender match" condition), the pronoun agreed in gender with the subject of the matrix clause. In the second version (the "gender mismatch" condition), the subject and pronoun disagreed in gender. Additionally, we constructed 90 sentence triplets. One version of each pair was a well-formed simple active sentence; a second version was grammatically well formed but contained a semantically inappropriate word (as determined by experimenter judgment; Table 2). A third, well-formed version of each sentence contained a reduced relative clause. (The responses to the relative clause sentences are not reported here.) A Latin-square design was used to form six stimulus lists from these materials, such that only one version of each experimental sentence appeared on each list, and with 30 exemplars of each type of sentence appearing on each list. An additional 30 grammatical filler sentences of various constructions were also added to each list. Hence, each subject was presented with 180 sentences. Ninety of these sentences were clearly well formed, whereas 30 of them were clearly anomalous. Sixty sentences (including 30 sentences containing reduced relative clauses and 30 sentences containing a subject noun that mismatched with a pronoun in gender) were technically well formed but were often judged to be unacceptable. The experimental and filler sentences were pseudorandomly mixed prior to presentation. The entire set of experimental sentences is provided in Appendix 2.

Procedure and Data Analysis

All procedures and data analyses were performed as in Experiment 1.

Results

Acceptability Judgments

Subjects judged the "gender match" and "gender mismatch" sentences to be acceptable

on 90 and 24% of the trials, respectively ($F(1,11) = 88.65, p < .0001$). For the semantically anomalous sentences and non-anomalous control sentences, the percentages were 9 and 80, respectively ($F(1,11) = 195, p < .0001$).

Event-Related Potentials

ERPs to sentences containing pronouns.

Across all subjects, 20% of the trials involving words embedded within sentences were rejected for artifact, while 26% of the trials involving sentence-final words were rejected. ERPs (representing grand averages over all subjects and trials) to the pronouns in the "gender match" and "gender mismatch" sentences are shown in Fig. 5. No reliable differences between conditions were found from 50 to 300 ms. Within the 300–500 ms window, ERPs to "mismatching" pronouns were more positive-going at Pz and Cz but more negative-going at Fz (sentence-type \times electrode site interaction, $F(2,22) = 4.23, p < .05, MS_e = 2.52$). Similar differences were observed at lateral sites ($F(4,44) = 4.55, p < .05, MS_e = 2.23$). The most notable differences between conditions were found in the 500–800 ms window, in which ERPs to mismatching pronouns were more positive-going than those to matching pronouns. However, this effect was smaller in amplitude than that seen in Experiment 1 and was largely restricted to posterior sites (sentence type \times electrode site interaction: midline, $F(2,22) = 4.26, p < .05, MS_e = 2.59$).

As noted above, the binding constraints of English grammar do not require pronouns to be bound to an explicitly mentioned antecedent within the sentence. Hence, subjects were free to assume that the pronoun referred to some unmentioned entity, thereby avoiding an apparent gender agreement violation. To determine whether the ERPs elicited in response to pronouns in the gender mismatch condition were a function of subjects' proclivity for interpreting the pronoun and subject to be coreferential, we reexamined subjects' acceptability judgments to the "gender mismatch" sentences. Subjects were then divided into two groups based on these judgments. One group ($N = 8$)

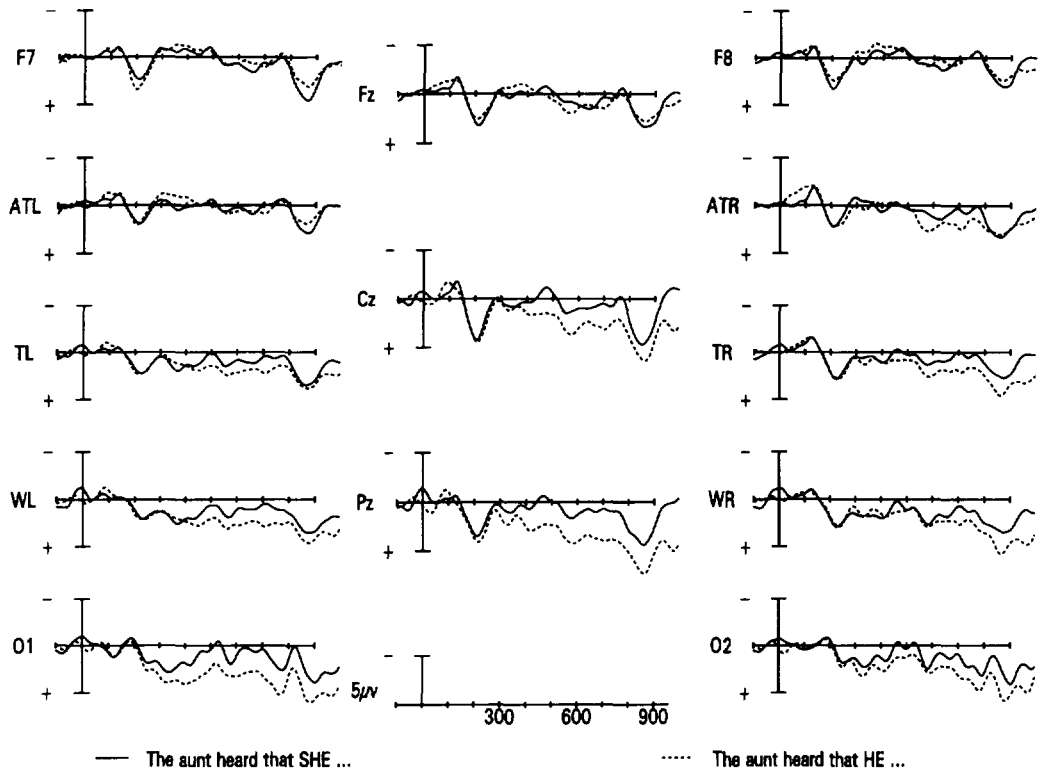


FIG. 5. Grand average ERPs (over all 12 subjects) to pronouns that matched (solid line) or mismatched (dashed line) with the subject noun in gender.

was comprised of subjects who rarely judged the "gender mismatch" sentences to be acceptable (range: 0 to 20%; mean: 7%). A second group of subjects ($N = 4$) often judged these sentences to be acceptable (range: 43 to 90%; mean: 63%).³

ERPs to pronouns for the group that typically judged the gender mismatch sentences to

³ An alternative procedure for computing response-contingent averages would be to sort responses on a trial-by-trial basis; e.g., one could average over only those trials on which the subject responded "acceptable." The problem is that this approach reduces the number of trials comprising each subject average, thereby reducing the signal-to-noise ratio for each average beyond acceptable limits. Also, we should note that although we cannot know for certain that subjects' acceptability judgments in Experiment 2 were completely determined by whether or not they took the subject noun and pronoun to be coreferential, it seems plausible to argue that this was true on most trials for most subjects, since the sentences were well-formed and coherent in every other respect.

be unacceptable are plotted in Fig. 6. Within the 500–800 ms window, ERPs to mismatching pronouns were reliably more positive-going than controls (midline: $F(1,7) = 16.45$, $p < .01$, $MS_e = 9.77$; lateral: $F(1,7) = 12.75$, $p < .02$, $MS_e = 15.16$). ERPs recorded from subjects who often judged the gender mismatch sentences to be acceptable are shown in Fig. 7. Although these waveforms should be interpreted cautiously due to the relatively few subjects comprising the grand average, it is clear that the "mismatching" pronouns did not elicit increased positivity within the 500 to 800 ms window. Statistical analysis found no reliable differences between conditions, $p > .2$ in all analyses.

Figure 8A plots ERPs to sentence-final words in the two sentence types. ERPs in the "gender mismatch" condition were reliably more negative-going than those in the gender match condition, beginning early in the epoch

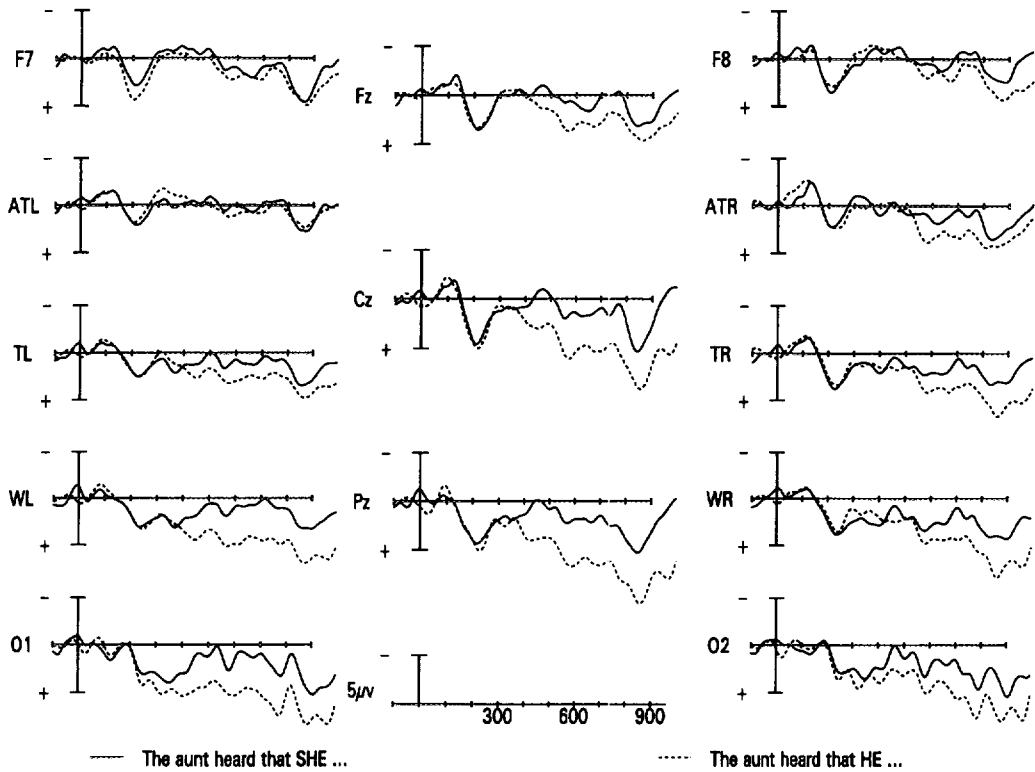


FIG. 6. Grand average ERPs to pronouns that matched (solid line) or mismatch (dashed line) with the subject noun in gender, averaged over the eight subjects who typically judged the mismatching sentences to be unacceptable.

at most sites. Reliable differences were found within the 300–500 ms (midline, $F(1,11) = 6.94$, $p < .05$, $MS_e = 22.62$; lateral, $F(1,11) = 8.81$, $p < .05$, $MS_e = 31.64$) and 500–800 ms windows (midline, $F(1,11) = 8.29$, $p < .05$, $MS_e = 16.13$; lateral, $F(1,11) = 10.04$, $p < .05$, $MS_e = 18.50$). Figures 8B and 8C plot ERPs to sentence-final words for subjects who typically judged the gender mismatch sentences to be unacceptable and acceptable, respectively. For the “unacceptable” group, ERPs to final words in gender mismatch sentences were more negative-going within 300–500 ms (midline: $F(1,7) = 10.24$, $p < .05$, $MS_e = 21.62$; lateral: $F(1,7) = 9.85$, $p < .05$, $MS_e = 44.97$) and 500–800 ms windows (midline: $F(1,7) = 11.88$, $p < .05$, $MS_e = 26.22$; lateral: $F(1,7) = 12.77$, $p < .01$, $MS_e = 10.85$). Furthermore, these differences between conditions tended to be posteriorly distributed (sentence type \times

electrode site, 300–500 ms window; midline: $F(1,7) = 5.61$, $p < .05$, $MS_e = 2.05$; lateral: $F(1,7) = 8.44$, $p < .05$, $MS_e = 6.55$; 500–800 ms window, lateral: $F(1,7) = 14.33$, $p < .01$, $MS_e = 2.95$). For the “acceptable” group, no reliable differences were found at midline sites within any window. However, ERPs to gender mismatch sentences were more negative-going than controls at some anterior lateral sites (500–800 ms window: $F(4,12) = 3.66$, $p = .05$, $MS_e = 1.58$).

ERPs to sentences containing semantically anomalous words. ERPs to semantically inappropriate words and controls are shown in Fig. 9. (Nineteen percent of the trials were rejected for artifact in each sentence type condition.) Semantically anomalous words elicited a greatly enhanced N400 component between 300 and 500 ms (midline: $F(1,11) = 11.46$, $p < .01$, $MS_e = 13.13$; lateral: $F(1,11) = 7.04$, $p < .05$,

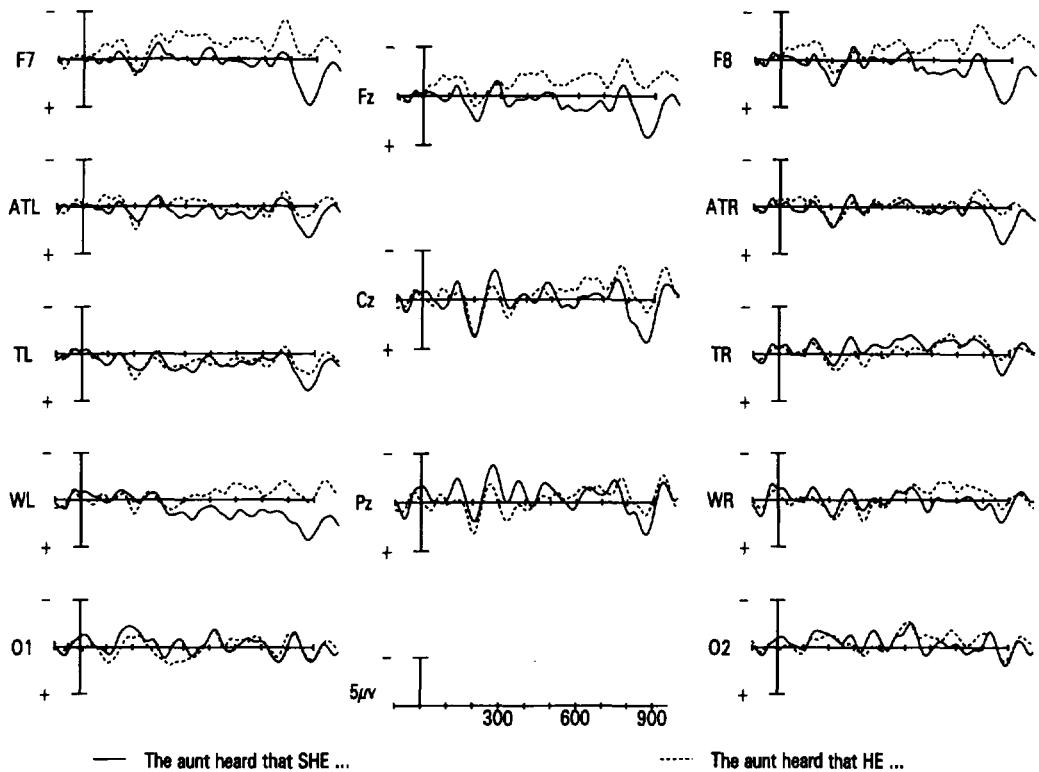


FIG. 7. Grand average ERPs to pronouns that matched (solid line) or mismatched (dashed line) with the subject noun in gender, averaged over the four subjects who typically judged the mismatching sentences to be acceptable.

$MS_e = 21.63$). This was followed by a positive-going component (relative to controls) at several lateral electrode sites (e.g., TR) beginning at about 600 ms. However, no reliable differences between conditions were observed within the 500–800 ms window, $p > .2$ in all analyses. ERPs to sentence-final words in sentences containing semantically anomalous words and control sentences are shown in Fig. 10. The final words in sentences containing anomalous words elicited an enhanced N400-like effect between 300 and 500 ms (midline: $F(1,11) = 5.80$, $p < .05$, $MS_e = 30.83$; lateral: $F(1,11) = 4.51$, $p = .05$, $MS_e = 34.98$).

Discussion

Pronouns that disagreed with the subject noun in gender elicited a P600-like positivity, but this was true only for subjects who typically judged these sentences to be unacceptable.

This finding indicates that the P600 response observed to agreement-violating reflexives in Experiment 1 was not elicited by a perceived violation of binding constraints. As in many previous experiments, semantically anomalous words elicited an enhanced N400 component rather than a P600-like positivity. This effect was observed even though subjects were asked to make sentence-acceptability judgments. Hence, the P600 appears not to be a general response to any type of unexpected, task-relevant linguistic anomaly under conditions in which subjects make sentence-acceptability judgments (see also Neville et al., 1991; Osterhout, 1990; Osterhout, Nicol, McKinnon, Fodor, Ni, & Crain, 1994).

EXPERIMENT 3

The results of Experiments 1 and 2 indicate that under conditions in which subjects make

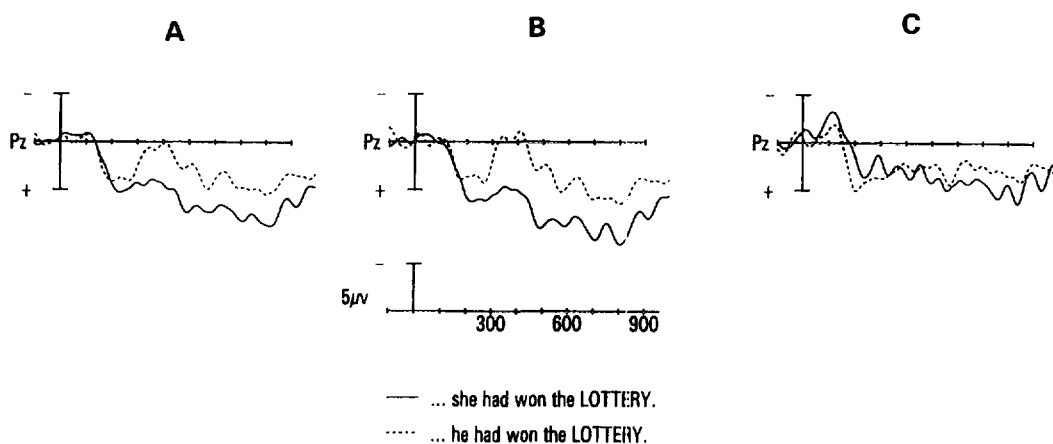


FIG. 8. Grand average ERPs (recorded over site Pz) to sentence-final words in sentences that contained pronouns that matched (solid line) or mismatched (dashed line) with the subject noun in gender. (A) ERPs recorded from all 12 subjects. (B) ERPs recorded from eight subjects who typically judged the "mismatch" sentences to be unacceptable. (C) ERPs recorded from four subjects who often judged the "mismatch" sentences to be acceptable.

sentence-acceptability judgments, several types of agreement violations elicit a positive-going wave that is quite distinct from the N400 effect elicited by semantically anomalous words. However, there are two problematic aspects of the design of these experiments. First, it is conceivable that the differential response to the two categories of anomaly is (at least partly) an artifact of the sentence-acceptability judgment. That is, perhaps these ERP effects reflect processes that are recruited in order to make the acceptability judgments but that are not routinely employed during "normal" reading. In Experiment 3, subjects were asked to read the sentences for comprehension and did not perform any additional task. The second problem concerns the composition of the stimulus lists presented in Experiment 1. Specifically, a relatively high proportion (45%) of the sentences presented during Experiment 1 contained agreement violations, and almost all of the anomalies were violations of agreement. It could be that the salience of the agreement violations influenced the detection of and/or the response to these anomalies.⁴ Therefore, in Ex-

periment 3 we manipulated the proportion of agreement violations subjects encountered during the experiment.

Method

Subjects

Sixteen undergraduate and graduate students (8 females and 8 males) participated for course credit or for a small compensation. None of the subjects participated in Experiments 1 or 2. Ages ranged from 18 to 35 (mean = 21) years.

Materials

Two sets of stimulus lists were prepared, to be presented in a between-subjects manner. The first set of lists was identical to the set presented during Experiment 1. For the second set, two sets of 60 sentence pairs that had been presented during Experiment 1 (the subject-verb number and reflexive-antecedent number sentences) were included. Also included were 60 sentence pairs containing semantically anomalous words or non-anomalous control words from Experiment 2. These materials were counterbalanced across two stimulus lists in the manner described for Experiment 1. Thirty reduced relative clause sentences from Experiment 2 and the set of 30 filler sentences from Experiment 1 were then

⁴ This interpretation runs contrary to the claim that the P600 is a member of the P300 family of components, since P300 amplitude is inversely related to the probability-of-occurrence of the anomaly; that is, the more probable the anomaly, the smaller the amplitude of the P300. In contrast, the salience argument predicts the opposite effect.

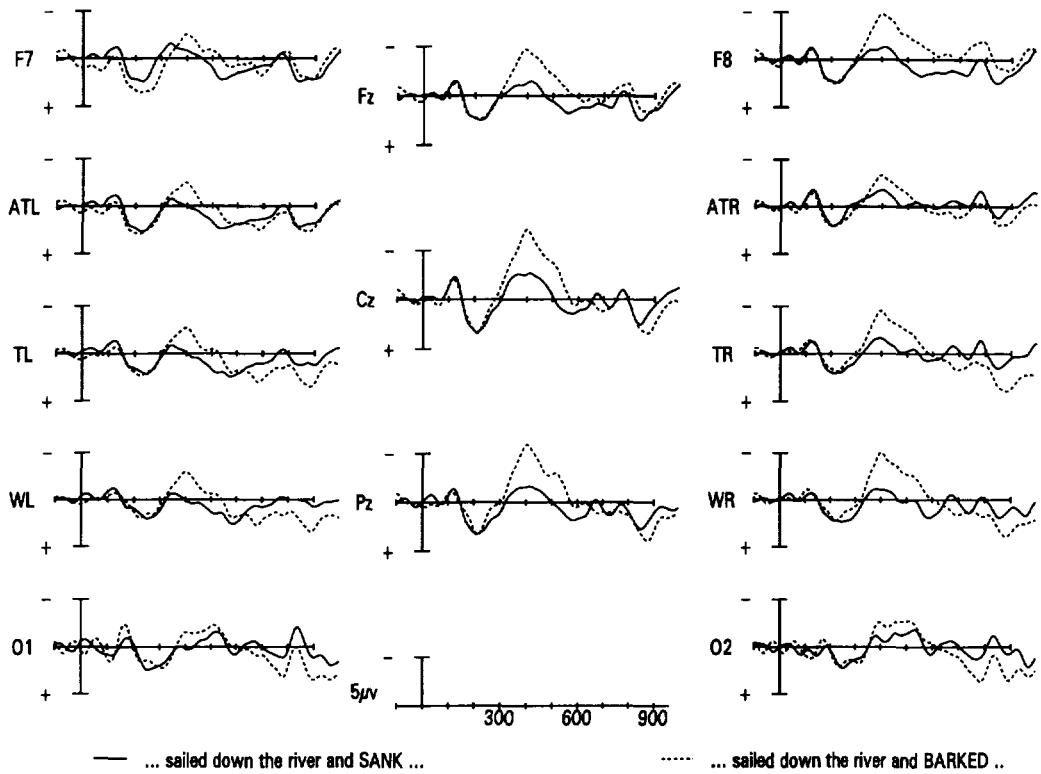


FIG. 9. Grand average ERPs to semantically anomalous words (dashed line) and non-anomalous control words (solid line), Experiment 2.

added to each list. Thus, subjects who saw one of the lists from the second set saw a total of 240 sentences, 105 of which were anomalous and 30 of which were technically well formed but typically judged to be unacceptable by readers. In Set 1, 45% of the sentences contained agreement violations; only 27% of the

sentences in Set 2 contained such violations. The percentage of anomalies in each set that involved agreement were 90 and 62%, respectively.

Procedure

These were identical to the procedures employed during Experiments 1 and 2, with one change. Subjects were asked to carefully read each sentence for comprehension. No other task was assigned. Subjects were also told to press a button on the joystick whenever they were ready for the next sentence when prompted with the message on the CRT screen to make a response.

Results

ERPs to Subject-Verb Number and Reflexive-Antecedent Number Disagreement

Because both groups of subjects saw the same sets of violating and control pairs in the

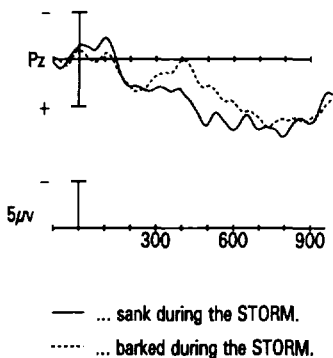


FIG. 10. Grand average ERPs (recorded over site P_z) to sentence-final words in sentences that contained a semantically anomalous word (dashed line) and well-formed control sentences (solid line), Experiment 2.

subject-verb number and reflexive-antecedent number conditions, ERPs in these conditions were combined for purposes of data analysis. Approximately 11% of the trials, roughly distributed equally across treatment conditions, were rejected due to artifact. Data from midline sites were analyzed in a three-way ANOVA with a between-subjects factor of stimulus set (set 1 or set 2) and within-subject factors of sentence type and electrode site. Data from lateral sites were analyzed in a four-way ANOVA with the added within-subject factor of hemisphere. In these analyses, the effects of list content could be evaluated by examining the interaction between stimulus set and sentence type.

ERPs to the subject-verb agreement violations and well-formed controls, from all 16 subjects, are shown in Fig. 11. Small differences between conditions were evident early in the epoch. However, the most striking difference was a positive shift in the ERPs to agree-

ment-violating verbs, relative to the non-violating controls, that had an onset around 500 ms. This effect was quite similar to the response elicited by the same words in Experiment 1, although somewhat lesser in amplitude. ANOVAs on mean amplitude within the 50–150, 150–300, and 300–500 ms windows revealed only one reliable effect, an interaction at midline sites between sentence type and electrode site between 50 and 150 ms, $F(2,28) = 4.31$, $p < .05$, $MS_e = 1$. This effect reflected the fact that the N100 component elicited by agreement-violating verbs was slightly larger than that elicited by controls, but only at Pz. Within the 500–800 ms window, ERPs to agreement-violating verbs were more positive-going than controls, both at midline, $F(1,14) = 5.04$, $p < .05$, $MS_e = 31.13$, and at lateral sites, $F(1,14) = 5.08$, $p < .05$, $MS_e = 45.89$. Furthermore, the interaction between stimulus set and sentence type was highly unreliable, both in the midline analysis, $p > .95$, and in the lat-

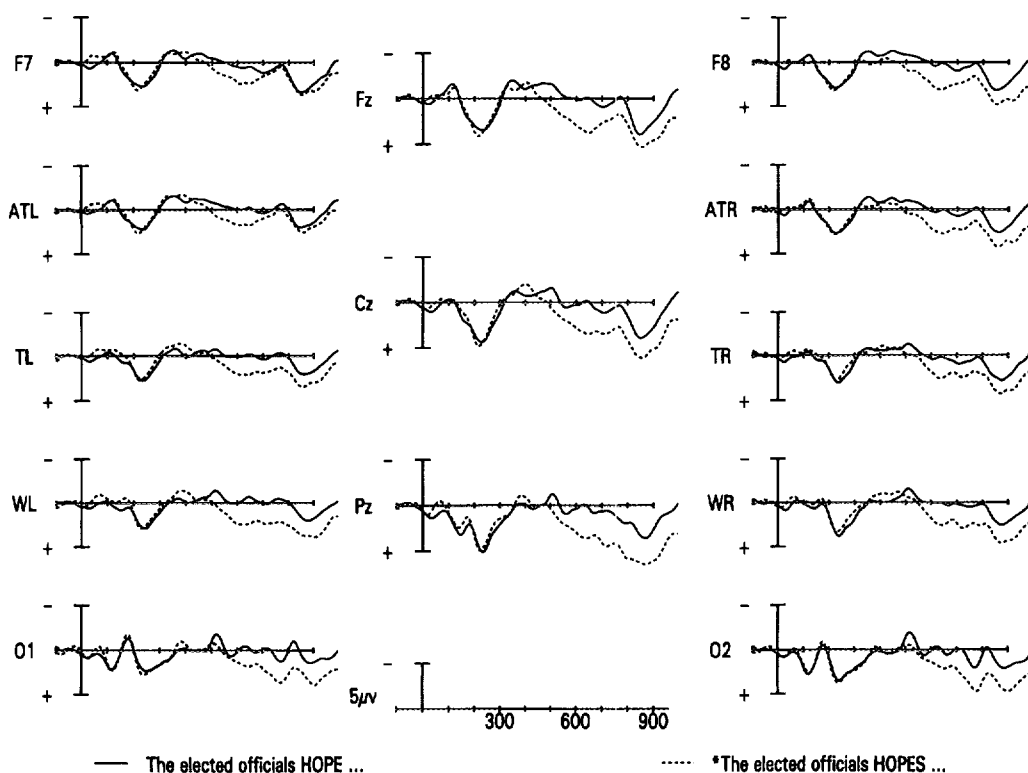


FIG. 11. Grand average ERPs to subject-verb number agreement violations (dashed lines) and controls (solid line), Experiment 3 ($N = 16$).

eral site analysis, $p > .90$. This indicates that list composition had negligible effects on the response to these violations.

ERPs to reflexive-antecedent number violations and controls are shown in Fig. 12. In contrast to the responses to these words observed during Experiment 1, only small differences between conditions were evident. ANOVAs on mean amplitude within these windows revealed that these differences were highly unreliable. Again, there were no reliable interactions between stimulus set and sentence type in any of the analyses.

ERPs to Reflexive-Antecedent Gender Disagreement and Semantic Anomalies

Because separate groups of subjects saw sentences containing reflexive-antecedent gender agreement violations and semantically anomalous words, data from these conditions were analyzed without the between-subjects factor of stimulus set. About 9% of the trials were re-

jected due to artifact, again roughly equally distributed across treatment conditions. ERPs to reflexive-antecedent gender disagreements (from eight subjects) are shown in Fig. 13. As was the case for the reflexive-antecedent number violations, and in contrast to Experiment 1, only small differences existed between violating reflexives and controls. ANOVAs found no reliable differences between conditions in any of the time windows.

ERPs to semantically anomalous words and controls are plotted in Fig. 14. As in Experiment 2, such words elicited a large increase in negativity, most notably between roughly 300 and 500 ms. A reliable main effect was found within this window, both at midline, $F(1,7) = 7.55$, $p < .05$, $MS_e = 9.58$, and at lateral sites, $F(1,7) = 5.62$, $p < .05$, $MS_e = 8.48$. Furthermore, differences between conditions were slightly larger in the left hemisphere, $F(1,7) = 5.39$, $p = .05$, $MS_e = 1$. Although these differences extended into the 500–800 ms win-

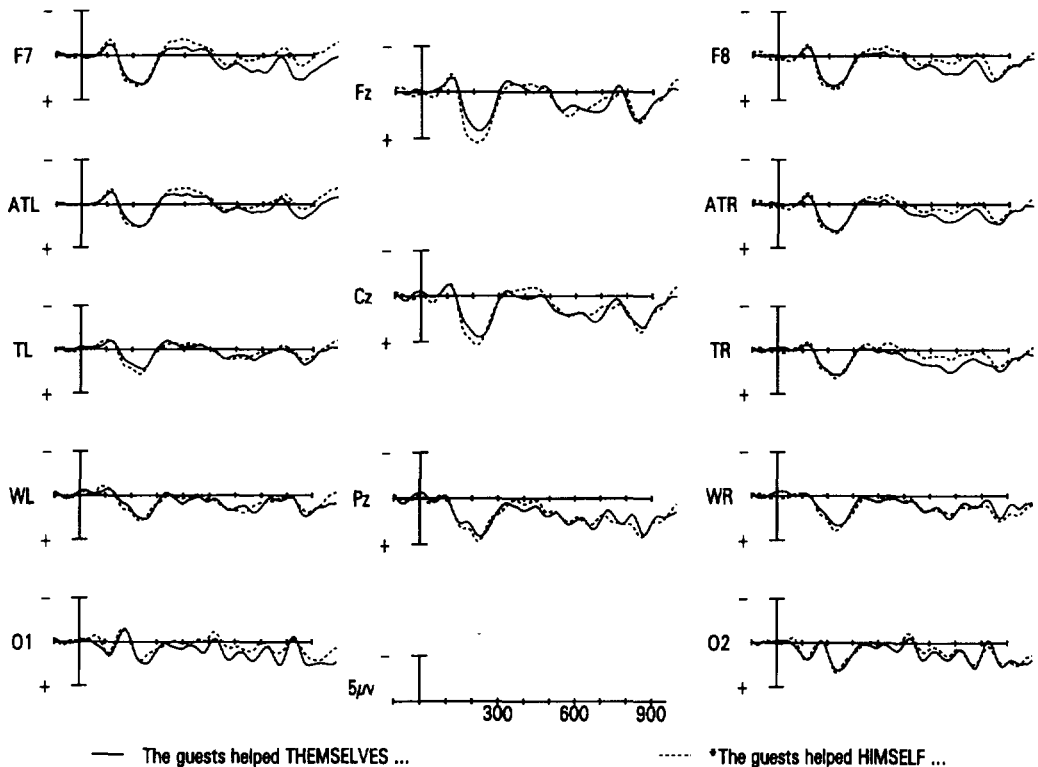


FIG. 12. Grand average ERPs to reflexive-antecedent number agreement violations (dashed line) and controls (solid line), Experiment 3 ($N = 16$).

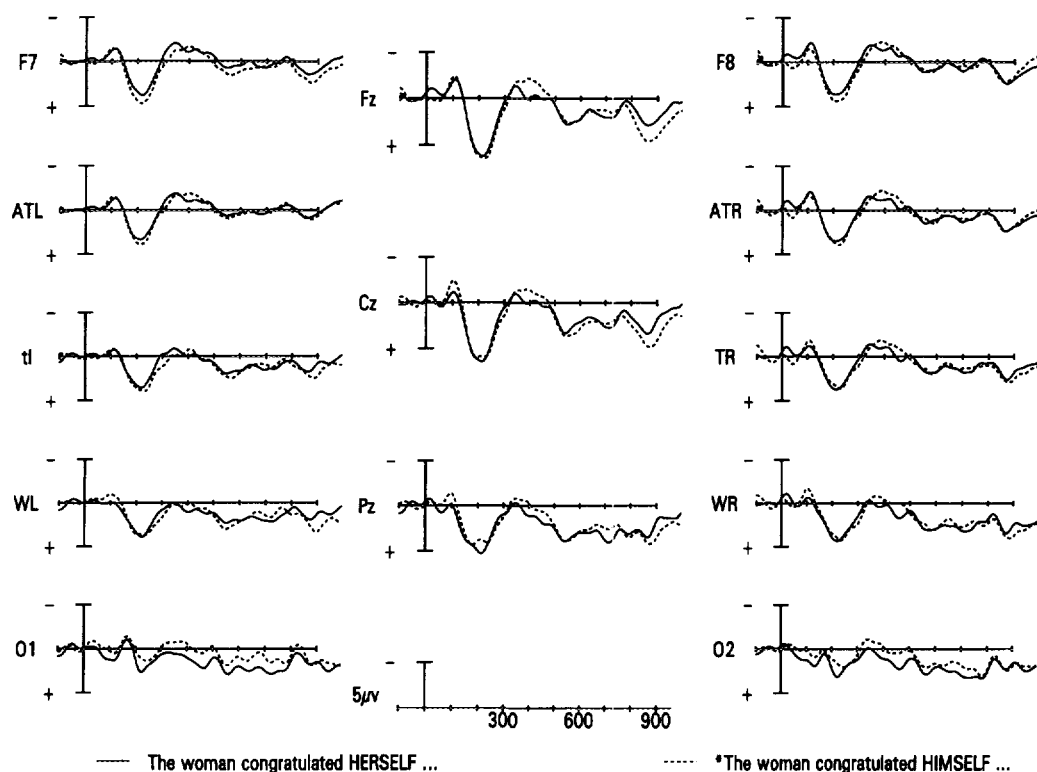


FIG. 13. Grand average ERPs to reflexive-antecedent gender agreement violations (dashed line) and controls (solid line), Experiment 3 ($N = 8$).

dow, no reliable differences were found in the later window.

ERPs to Sentence-Ending Words

Inspection of ERPs to sentence-final words revealed only small differences in the ERPs to sentences containing a violation and controls. The only exception was that ERPs to the sentence-final words in sentences containing a semantically anomalous word were noticeably more negative-going than controls, beginning at about 300 ms and continuing throughout the epoch. Furthermore, differences between conditions were largest posteriorly. ANOVAs on mean amplitude within the 300–500 ms window revealed a reliable interaction between sentence type and electrode site at midline sites, $F(2,14) = 14.11$, $p < .01$, $MS_e < 0.27$ and a main effect of sentence type at lateral sites, $F(1,7) = 5.38$, $p = .05$, $MS_e =$

11.10. Similar effects were obtained within the 500–800 ms window (sentence type \times electrode: midline, $F(2,14) = 14.67$, $p < .01$, $MS_e < 0.48$; main effect for sentence type: lateral, $F(1,7) = 5.91$, $p < .05$, $MS_e < 0.89$). No other analyses approached significance in any of the conditions.

Discussion

Under conditions in which subjects passively read sentences without performing a secondary task, subject–verb number agreement violations and semantically anomalous words elicited clearly distinct brain responses. These responses were quite similar to those observed when subjects performed sentence-acceptability judgments. Furthermore, the proportion of agreement violations within the list seen by subjects had no reliable effects on the responses to these violations. These findings indicate that the differential response to agreement viola-

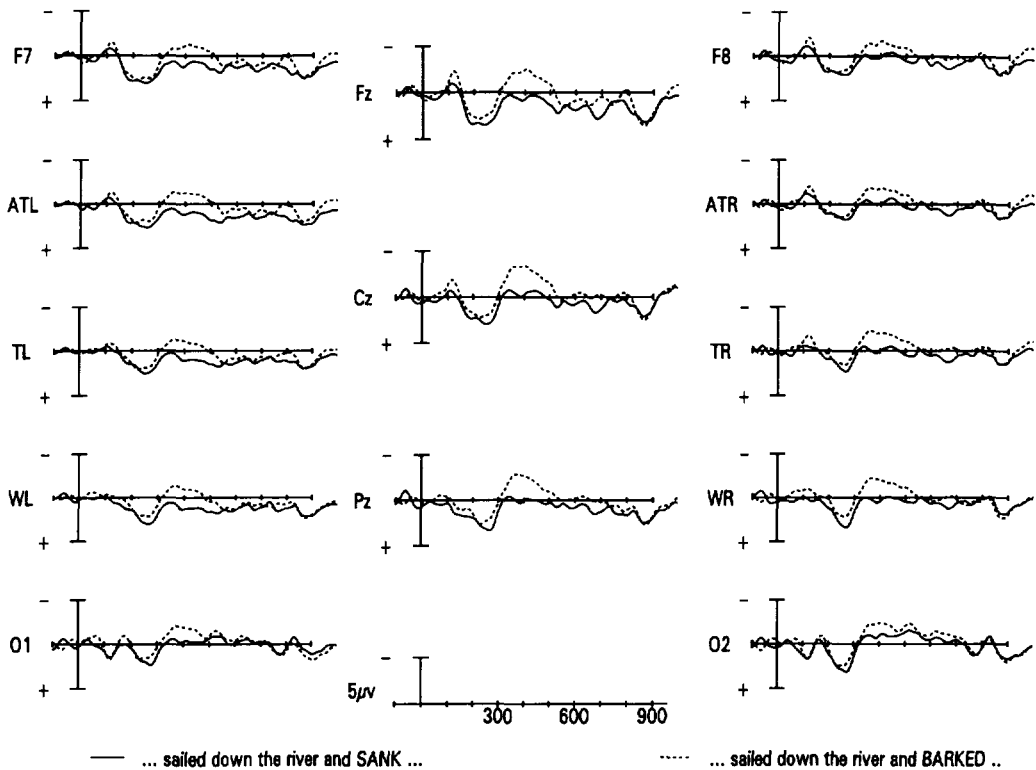


FIG. 14. Grand average ERPs to semantically anomalous words (dashed line) and controls (solid line), Experiment 3 ($N = 8$).

tions and semantic anomalies is not an artifact of the sentence-acceptability judgment task or of list content. However, the results of Experiment 3 differed from those of Experiments 1 and 2 in that reflexive–antecedent number and gender disagreement elicited a P600 when subjects were asked to make sentence-acceptability judgments (Experiment 1) but not when subjects passively read sentences (Experiment 3). By contrast, subject–verb number disagreement elicited a P600 under both task conditions. Thus, although the differential response to agreement and semantic anomalies was not a function of task, task factors seemed to influence whether or not ERPs to the violating reflexives were different from those to controls.

GENERAL DISCUSSION

In three experiments, we recorded ERPs while subjects read sentences, some of which contained an agreement violation or a semantically anomalous word. Subjects performed

sentence-acceptability judgments (Experiments 1 and 2) or read sentences without performing any additional task (Experiment 3). In all three experiments, violations of number and gender agreement (excepting reflexive–antecedent disagreements in Experiment 3) elicited a widely distributed positive-going wave. This effect resembled the positivities previously observed in the response to several types of syntactic anomaly (the “P600” effect; Hagoort et al., 1993; Neville et al., 1991; Osterhout, 1990; Osterhout & Holcomb, 1992; Osterhout & Holcomb, 1993; Osterhout et al., 1994). Semantically anomalous words elicited a negative-going wave (the “N400” effect) that was quite distinct from the P600 (see also Kutas & Hillyard, 1980a, 1980b, 1980c; Neville et al., 1991; Osterhout, 1990; Osterhout, Nicol, et al., 1994). Assuming that similar patterns of brain activity reflect similar sets of neural and cognitive events, whereas the converse is true of dissimilar patterns of brain activity, one could in-

interpret these results as indicating that violations of agreement are perceived to be syntactic, rather than semantic, in nature.

There are, however, several objections to this conclusion. One objection concerns the potential effects of component overlap. Specifically, the large positivity elicited by agreement-violating words might have obscured a robust N400 effect. Similarly, the large N400 to semantically anomalous words might have obscured a robust P300/P600. These possibilities gain some credence upon noting that in some cases agreement violations elicited slightly larger "N400s" at some sites, whereas semantically anomalous words in Experiment 2 elicited slightly increased positivity within the "P600 window" at several sites. According to this view, the responses to the two categories of anomaly represent differential modulations of an (at least partially) overlapping set of neural systems. The problem of overlapping components is such that this possibility can be neither discounted nor confirmed, given the data in hand. However, our claim is that at least under the conditions of the present study, the response to a variety of syntactic anomalies and agreement violations is dominated by a positive-going wave, whereas the response to semantically anomalous words is dominated by a negative-going wave. Even if the agreement violations elicited an increase in N400 amplitude, that increase was dwarfed both by the much larger positivity and by the magnitude of the N400 effect elicited by semantically anomalous words. Conversely, even if the semantic anomalies elicited a P600-like positivity, that effect was dwarfed both by the greatly enhanced N400 component and by the P600 elicited by the agreement violations. Hence, responses to the two categories of anomaly are clearly distinct.

Another objection is that these results might not generalize to other comprehension environments. A particular worry is that the effects of anomaly type have been confounded with the effects of task. One salient hypothesis in this regard is that the P600 to the agreement violations is a member of the P300 family of positivities known to be elicited by unexpected, task-relevant stimuli (Donchin, 1981; Dun-

can-Johnson & Donchin, 1977). Presumably, the agreement violations were task relevant both when subjects made sentence-acceptability judgments and when they passively read the sentences. However, this interpretation must somehow account for the absence of a robust P300/P600-like effect in the response to semantically anomalous words (Experiments 2 and 3), which were also task relevant and unexpected (see also Neville et al., 1991; Osterhout, 1990; Osterhout, Nicol, McKinnon, Ni, & Fodor, 1994). Consider also a recent study reported by Osterhout, Nicol, et al. (1994), who directly contrasted the response to well-formed control sentences (e.g., "The cat won't *eat* the food . . ."), inflectional violations ("The cat won't *eating* the food . . ."), and (semantic) selectional restriction violations ("The cat won't *bake* the food . . ."). The inflectional anomalies elicited a large P600-like effect, whereas the selectional restriction violation elicited a monophasic increase in N400 amplitude. Finally, recent work in our laboratory has indicated that the P300 and P600 are to an interesting extent independent effects with distinct morphologies, time courses, and scalp distributions, and that the two effects have differential sensitivities to changes in task and stimuli (Osterhout, McKinnon, Bersick, & Corey, 1995). Thus, it seems unlikely that the P600 response to syntactically anomalous words is a general response to the unexpectedness or task-relevance of *any type* of linguistic anomaly. Furthermore, the observation that agreement violations and semantically anomalous words elicited the P600 and N400 effects under two distinct task conditions (sentence-acceptability judgments and passive reading) suggests that the differential response to these anomalies is not simply a task artifact (see also Hagoort et al., 1993).

We are not claiming that the choice of task had *no* effect on the ERP response to anomalies. Indeed, task effects are the most likely explanation for the observation that reflexive-antecedent agreement violations elicited a P600 effect when subjects made sentence-acceptability judgments but not when subjects passively read the sentences. Our speculation is that the task influenced the likelihood of

anomaly detection. That is, perhaps subjects rapidly and reliably computed the coreferential relations between reflexives and antecedents when making sentence-acceptability judgments but not when passively reading the sentences. Several researchers have suggested that the time course of coreferential computations might depend both on elements of the discourse context (Greene, McKoon, & Ratcliff, 1992; Sanford & Garrod, 1989) and on the nature of the task given to subjects (Oakham, Graham, & Vonk, 1989). According to this view, coreferential relations are not always computed immediately (Green et al., 1992; Sanford & Garrod, 1989) or, in some situations, are not computed at all (Greene et al., 1992; Oakham et al., 1989). In the present study, determining the referent to the reflexive was clearly crucial for successful completion of the sentence acceptability judgment but was not explicitly necessitated by the passive reading task (Oakham et al., 1989). The effects of task might have been amplified by the nature of the stimuli (Greene et al., 1992), in particular the repetitive presentation of unrelated sentences, most of which introduced a new subject with its own number and gender properties. Given the signal averaging procedures used here, a small or null effect would result if subjects did not consistently compute the coreferential relations or if there was considerable temporal variation in the coreferential processes. Regardless of the proper interpretation of this task effect, the critical point is that although the change in task appeared to determine whether the reflexive agreement anomalies elicited a response that differed from that to well-formed controls (i.e., whether or not such anomalies were rapidly and reliably detected), there is no evidence in the current study that any of the anomalies elicited qualitatively different responses under different task conditions. Even so, we do not know whether the conclusions reached here will extend to more "natural" environments, e.g., situations in which subjects read connected discourse for comprehension with no additional task assigned. In our laboratory we are currently examining this important question by systematically observing the effects of task and presentation mode on the response to a va-

riety of linguistic anomalies (including agreement violations).

We should note that the existing literature reveals limitations to the generalization that syntactic and semantic anomalies elicit the P600 and N400 effects, respectively. Recent studies have reported that certain types of "syntactic" anomalies elicit a negative-going response rather than (or in addition to) a late positivity (Friederici, Pfeifer, & Hahne, *in press*; Munte, Heinze, & Mangun, 1993; Rosler, Friederici, Putz, & Hahne, 1993), at least under certain conditions. Although the distinct scalp distributions of the negativities elicited by "syntactic" and "semantic" anomalies have permitted researchers to maintain the claim that these two categories elicit distinct brain responses, such a finding is clearly inconsistent with the claim that syntactic anomalies elicit the P600 response under all experimental conditions. Similarly, there are reported instances in which sentence-final semantically anomalous words elicit both an N400-like effect and a subsequent robust positive-going wave (e.g., Kutas, 1993; Neville, Mills, & Lawson, 1992). One explanation for these discrepancies hinges on the possibility that the ERP response to a linguistic anomaly is influenced by the anomalous word's sentence position (e.g., sentence-final vs. sentence-embedded). In particular, by presenting the anomalous word in sentence-final position, one risks confounding the response to the anomaly with sentence wrap-up, decision, and response factors (for more discussion see Osterhout, 1994, or Osterhout & Holcomb, 1995). The primary claim of the present study is that *under the experimental conditions employed here*, syntactic and semantic anomalies elicit distinct brain responses, thereby providing a diagnostic tool for determining the level of analysis at which some comprehension phenomenon is perceived to occur.

A final set of objections concerns the claim that pronouns that mismatch in gender with the subject noun are perceived to be syntactically anomalous. These "mismatching" pronouns violated no grammatical rule. How, then, can we maintain our claim that the P600 is associated with syntactic violations? We believe the answer to this objection resides in the im-

portant distinction between *outright* violations of a formal rule and *perceived* violations that result from the operations of the comprehension system. This distinction is clearest in the large literature on the comprehension of sentences containing local syntactic ambiguity, i.e., situations in which two syntactic analyses can be assigned to a single sentence fragment. A substantial body of evidence indicates that subjects pursue a single "preferred" syntactic analysis in such situations. If subjects pursue the analysis that eventually turns out to be the wrong one, the disambiguating portion of the sentence cannot be attached to the parse tree, and the sentence is perceived (at least momentarily) to be ungrammatical (Frazier & Rayner, 1982; Osterhout et al., 1994; Rayner, Carlson, & Frazier, 1983; for review, Frazier, 1987). *Under the analysis pursued by the subject*, the sentence is indeed ill formed. We believe that the results we report in Experiment 2 are analogous to the large literature on the parsing of syntactically ambiguous sentences. What we have observed in Experiment 2 is a form of the garden-path effect that involves coreference. If subjects take the pronoun and the subject noun to be coreferential, then this interpretation violates the presumed formal requirement that certain coreferring elements, including pronouns and their antecedents, must agree in number and gender.

The conclusion that the "mismatching" pronouns were perceived to be syntactically anomalous also appears to contradict recent claims that pronoun-antecedent relations are computed at the conceptual level. For example, one recent theory posits a discourse preference rule stating that pronouns must be interpreted in relation to previously presented information (Gordon, Grosz, & Gilliom, 1993). This would lead to the coreferencing of *he* and *the actress* in the sentence *The actress predicted that he would get the job*. However, the present conclusions do not necessarily contradict this discourse rule. Rather than reflecting the (possibly semantic/conceptual) factors that determine which elements are taken to be coreferential, the P600 might reflect whether or not the (presumably syntactic) agreement features of the

coreferring elements match. Less easily reconciled is evidence that conceptual properties of the controller, under certain circumstances, determine the preferred agreement features of the controllee. For example, although American English speakers typically use singular verb agreement with collective nouns (e.g., *committee*), they often use plural pronoun agreement with such nouns (Bock & Eberhard, 1993). One possibility is that collective nouns represent a special category for which semantic and conceptual factors play a larger role in agreement phenomena. This hypothesis can be easily tested by recording ERPs to sentences containing collective nouns that act as controllers for verbs and pronouns.

It is important to point out that we do not know for certain precisely what aspect of the anomaly is eliciting the brain response to agreement violations. The P600 might reflect a realization that the grammatical rule requiring agreement between the controller and controllee has been violated. But this possibility runs into difficulty if pronoun-antecedent relations are determined by discourse factors. Another possibility, as noted above, is that the P600 was elicited in response to the mismatching features themselves. Even though features such as number and gender (and perhaps also person, definiteness, mass/count, etc) might have semantic bases (e.g., Corbett, 1991; Reid, 1991), such an interpretation implies that these features become "grammaticized" by virtue of the fact that they enter into agreement relations. The features might then become in a sense dissociable from the lexical item itself (specifically, from the lexical item's meaning). Indeed, any theory which requires the "percolation" of agreement features to higher nodes in a syntactic tree (as GB theory requires agreement features to percolate up to the noun phrase node from nouns within the sentence) assumes that such features are separable from the noun itself.

These conclusions are perhaps most striking as they apply to the determination of gender agreement between anaphoric entities (reflexives and pronouns) and their antecedents. In English, gender distinctions largely correspond

to the semantic distinctions between male, female, and neuter (Corbett, 1991). Furthermore, gender distinctions in English (unlike other languages) are not marked through such overt "grammatical" devices such as affixation (excepting the largely unproductive "feminine" suffixes "-ess" and "-ette"); nor do they enter into subject-verb agreement. Regardless, violations of gender agreement elicited a brain response similar to that elicited by a variety of syntactic anomalies, including violations of the more overtly "grammaticized" feature of number.⁵

Our speculation that agreement is represented as a formal constraint on well-formedness is consistent with recent claims stemming from examinations of subject-verb agreement errors induced during sentence production (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991). The major finding of these experiments has been that the probability of agreement errors during language production is strongly influenced by structural factors (e.g., errors seem to be clause-bound, and agreement between a subject noun and a matrix verb seems to be calculated with respect to abstract designations of subject) but relatively uninfluenced by semantic factors (e.g., semantic properties often associated with subjects, such as animacy, fail, by themselves, to engender agreement errors). Such evidence has been interpreted as indicating that structural relations are a necessary part of the agreement process while semantic relations are not (Bock & Miller, 1991). Furthermore, Bock and Eberhard (1993) recently examined the rate of verb number errors following nouns that were both "lexically" and "notionally" singular (e.g., *soldier*) and following nouns that were lexically singular but notionally plural (e.g., collective nouns such as *army*). Subjects did not produce more verbs in their plural form following the notionally plural nouns than following the notionally singular nouns, indicating that notional plurality does not have a strong influence on subject-verb agreement.

⁵ We thank Janet Nicol for suggesting the substance of the preceding two paragraphs.

In addition to the large P600 effects, small amplitude, early occurring differences between agreement-violating and control sentences were also observed. Intriguingly, these differences were not constant across agreement conditions. Most notably, subject-verb number agreement violations in Experiment 1 elicited both a left-hemisphere negativity between 300 and 500 ms and a small increase in P2 amplitude. Although these effects were not observed in Experiment 3, similar left-hemisphere effects have been observed following violations of phrase structure constraints under conditions similar to those used here (Neville et al., 1991; Osterhout & Holcomb, 1992, 1993). At present, it is unclear why subject-verb number agreement (and phrase structure violations) would elicit this effect while other types of "syntactic" violations (e.g., violations of verb subcategorization, subadjacency, and reflexive number and gender agreement) do not. If one assumes that the onset of reliably diverging waves temporally marks the discrimination of well-formed and ill-formed structures, then this finding would seem to indicate that subject-verb number agreement is detected more rapidly than agreement involving reflexive pronouns. However, the existence of small (and unreliable) differences between reflexive-antecedent agreement violations and controls, particularly in the N400 region, lends a cautionary note to such conclusions. Additional work, specifically designed to examine small-amplitude effects, is needed to examine the reliability and functional significance of these effects.

A final issue concerns the importance of generalizing the observed results beyond the particular set of items used in the experiment. A standard procedure within psycholinguistics for generalizing across items has been to perform analyses treating items as a random factor (cf. Clark, 1973). However, for a number of reasons item analyses of this type are rarely performed on ERP data and were not performed in the current study. (For extensive discussions of the problems associated with such analyses, see Garnsey, 1993, or Osterhout, 1994.) One reason for this is related to the signal-to-noise issue inherent in the signal-aver-

aging procedures used to obtain the ERP. In the current study, such analyses would involve grand averages over between 8 and 16 waveforms (reflecting the number of subjects in the analysis), a number insufficient to obtain the necessary signal-to-noise ratio. Running the number of subjects required to derive sufficient signal-to-noise ratios is often deemed prohibitively expensive in terms of the use of resources. Instead, most ERP researchers rely on replications across different sets of items to determine the generalizability of the effects of interest. In our laboratory we have replicated all of the basic effects reported here in studies incorporating partially or entirely new sets of stimuli (Osterhout, 1995a, 1995b; Osterhout et al., 1995).

In sum, the results reported here provide a preliminary empirical basis for the claim that, at least under the experimental conditions used here, readers perceive a variety of agreement violations encountered during reading to be syntactic in nature. Such a result allows one to speculate that agreement is part of the form, rather than the meaning, of language. At the same time, we have shown that violations of yet another type of linguistic constraint, one that is often treated as syntactic within formal theories, elicit an ERP response dominated by a late positive-going wave (P600). In sharp contrast, semantically inappropriate words elicit an ERP response dominated by a negative-going wave (N400). These findings are consistent with previous speculations (Osterhout & Holcomb, 1992) that in addition to the clear advantages of on-line, continuous, and non-intrusive measurement, ERPs, perhaps uniquely among currently used methods, might also have the significant advantage of a sensitivity to the "representational levels" of language.

APPENDIX 1

Experimental Sentences Presented during Experiment 1

Subject-Verb Number Agreement Sentences

1. Most cats like/likes to play outside.
2. Few students really know/knows how to study for exams anymore.

3. The newspapers always hope/hopes that presidential campaigns are close contests.
4. Baby gorillas make/makes excellent pets.
5. Many doctors claim/claims that insurance is too high.
6. The South Pacific islands are/is excellent vacation destinations.
7. Modern office buildings often resemble/resembles sheets of glass.
8. The local universities are/is hoping to recruit more adult students.
9. Most dentists recommend/recommends brushing twice daily.
10. Software companies know/knows that the industry will grow rapidly.
11. Old baseball cards are/is now very valuable.
12. Department stores are/is in desperate need of new customers.
13. The new bank policies seem/seems to be working well.
14. The reluctant witnesses appear/appears to be holding up well.
15. The newly elected officials hope/hopes to balance the budget.
16. Mechanical pencils allow/allows you to write with precision.
17. Soda bottles are/is recycled in many states.
18. Europeans often dislike/dislikes American tourists.
19. The ski slopes in Austria are/is very challenging.
20. Many artists paint/paints with water colors.
21. Florida alligators like/likes to eat raw hamburger.
22. Polar bears live/lives at the North Pole.
23. Most books have/has extensive indexes.
24. Stamp collectors buy/buys limited edition stamps as investments.
25. Political candidates travel/travels all over the country looking for votes.
26. Movie directors make/makes more money than many actors.
27. Many dentists take/takes two month vacations in the summer.
28. African elephants live/lives in the jungle.
29. Old elevators have/has creaky doors.

30. American farmers grow/grows more corn than all other farmers combined.

31. Young fathers need/needs to spend lots of time with their children.

32. Few gardeners know/knows how to grow exotic flowers.

33. Board games have/has become popular in recent years.

34. Most houses in the northeast have/has steep roofs.

35. Many hurricanes start/starts in the Caribbean.

36. Few inventors have/has had the impact of Thomas Edison.

37. College interviews are/is a crucial part of the application process.

38. Most juries agree/agrees on a verdict within hours.

39. Many judges hope/hopes to become Supreme Court Justices.

40. Car keys have/has a way of getting lost easily.

41. Most kittens claw/claws the furniture just for fun.

42. Few lawyers donate/donates their time to the poor.

43. Hot liquids become/becomes a gas at specific temperatures.

44. Many memories fade/fades after a few years.

45. Most meteors burn/burns up before they reach earth.

46. Many magicians know/knows how to escape from a safe.

47. Daily newspapers print/prints all the news that's fit to print.

48. Spotted owls require/requires two thousand acres of forest to survive.

49. Denver omelets contain/contains green peppers and onions.

50. Green peas contain/contains many important nutrients.

51. Fraternity parties are/is often very noisy.

52. Most queens wear/wears a crown.

53. Dimes and quarters are/is needed for parking meters.

54. Country roads receive/receives less maintenance than major highways.

55. Road races attract/attracts runners of all abilities.

56. Dry summers result/results in an increased risk of fire.

57. Shopping malls have/has become increasingly popular.

58. Passenger trains are/is more comfortable than buses.

59. Most butchers cut/cuts meat using cleavers.

60. Men's ties change/changes with fashion trends.

Antecedent-Reflexive Number Agreement Sentences

1. The persistent children ingratiated themselves/himself with the train conductor.

2. The careless pedestrians found themselves/herself covered with mud.

3. The emotional sisters felt themselves/herself going sentimental.

4. The successful hunters cleaned themselves/himself after walking through the woods.

5. The hungry guests helped themselves/herself to the delicious meal.

6. The hungry chefs cooked dinner for themselves/himself after work.

7. The nervous actors calmed themselves/himself before going on stage.

8. The tired waitresses poured some coffee for themselves/herself after work.

9. The bitter employees prepared themselves/herself for the confrontation.

10. The talented tailors made themselves/himself some beautiful clothes.

11. The angry customers helped themselves/herself to the merchandise.

12. The rowdy children helped themselves/herself to the cupcakes.

13. The careless scientists hurt themselves/himself with the dangerous chemical.

14. The sleazy politicians believed themselves/himself to be above the law.

15. The treacherous generals placed themselves/himself on the throne.

16. The famous actresses prepared themselves/herself to face the crowd.

17. The shy choir girls forced themselves/herself to sing the part.

18. The lonely girls played themselves/herself a song.

19. The novice cowgirls surprised themselves/herself with success.

20. The eccentric spinsters enjoyed themselves/herself on the trip.

21. The clumsy clowns tripped over themselves/himself during the performance.

22. The respected mayors honored themselves/himself with a party.

23. The talented stars saw themselves/himself in the movie.

24. The aspiring Rockettes taught themselves/herself to dance.

25. The experienced midwives established themselves/herself in the community.

26. The winning contestants considered themselves/himself to be very lucky.

27. The bubbly cheerleaders sewed the costumes themselves/herself before the season began.

28. The old ladies gathered flowers for themselves/herself every morning.

29. The obsessive lawyers worked by themselves/himself for many hours each day.

30. The talented tailors fixed the problem themselves/himself before the plumber arrived.

31. The popular senator promised himself/themselves a vacation.

32. The industrious saleswoman congratulated herself/themselves for earning a bonus.

33. The grateful niece asked herself/themselves how she could repay her aunt.

34. The capable girlscout built herself/themselves a fire.

35. The rowdy maid calmed herself/themselves with a glass of wine.

36. The famous writer described herself/themselves in glowing terms.

37. The revered leader revealed himself/themselves to be a fraud.

38. The ragged beggar relieved himself/themselves on the street corner.

39. The clumsy officer hurt himself/themselves during the training session.

40. The school principal talked to himself/themselves after the embarrassing incident.

41. The homeless widow built herself/themselves a shelter.

42. The nimble grandmother sewed herself/themselves some new dresses.

43. The unappreciated musician sent herself/themselves a bouquet of roses.

44. The nervous doctor told himself/themselves to relax before the operation.

45. The reluctant witness prepared himself/themselves for the trial.

46. The busy housewife threw herself/themselves into the housework.

47. The cold girl prepared herself/themselves some cocoa.

48. The famous ballerina prepared herself/themselves for the performance.

49. The convicted murderer killed himself/themselves after the verdict was announced.

50. The enraged ballplayer calmed himself/themselves after the devastating loss.

51. The Olympic swimmer trained herself/themselves for the swim meet.

52. The heavyweight boxer hurt himself/themselves before the match.

53. The dangerous gangster turned himself/themselves in to the authorities.

54. The forest ranger readied himself/themselves for the storm.

55. The brawny logger helped himself/themselves to the hearty breakfast.

56. The dirty soldier cleaned himself/themselves at the lake.

57. The seasick sailor threw himself/themselves onto the bunk.

58. The weary trucker allowed himself/themselves a one-hour nap.

59. The greasy mechanic considered himself/themselves to be very handsome.

60. The tired nurse administered the injection herself/themselves without telling the doctor.

Antecedent-Reflexive Gender

Agreement Sentences

1. The lonely bachelor cooked dinner for himself/herself after work.

2. The anxious cowboy prepared himself/herself for the rodeo.

3. The overweight deacon refused himself/herself a doughnut.

4. The overbearing patriarch found himself/herself alone.
5. The devout priest crossed himself/herself at the altar.
6. The insecure son thought himself/herself unloved.
7. The wicked stepfather put himself/herself in charge.
8. The greedy duke bought himself/herself a new limousine.
9. The vain prince looked at himself/herself in the mirror.
10. The nervous groom checked himself/herself in the mirror.
11. The young husband found himself/herself without a job.
12. The harassed congressman fixed himself/herself a drink.
13. The insecure king commissioned a statue of himself/herself for the park.
14. The old man got himself/herself out of bed to face the day.
15. The lovestruck boy kissed himself/herself on the arm for the practice.
16. The kindly uncle enjoyed himself/herself at Christmas.
17. The lonely grandfather made himself/herself a cup of tea.
18. The overworked salesman got himself/herself a beer.
19. The stubborn nephew found himself/herself written out of the will.
20. The dutiful boy scout quizzed himself/herself for the test.
21. The loyal butler prepared himself/herself for a controversy.
22. The picky chairman did the work himself/herself during the meeting.
23. The tardy milkman found himself/herself out of a job.
24. The elderly gentleman fixed himself/herself up for the dance.
25. The grateful grandson bought himself/herself new clothes.
26. The desperate boyfriend told himself/herself to forgive the girl.
27. The disoriented policeman lost himself/herself in the crowd.
28. The confused brother wrote himself/herself a note.
29. The brave fireman pulled himself/herself from the flames.
30. The ambitious count pictured himself/herself as king.
31. The athletic girl taught herself/himself how to shoot baskets.
32. The successful woman congratulated herself/himself on her promotion.
33. The airsick stewardess poured herself/himself a glass of water.
34. The hungry waitress ordered herself/himself a burger.
35. The novice actress embarrassed herself/himself on stage.
36. The wealthy queen built herself/himself a castle.
37. The neglected wife bought herself/himself an anniversary present.
38. The gracious hostess introduced herself/himself to the guests.
39. The unhappy duchess killed herself/himself after the scandal.
40. The infamous princess looked at herself/himself in the mirror.
41. The calm bride prepared herself/himself for the wedding.
42. The high priestess anointed herself/himself with oil.
43. The industrious daughter put herself/himself through school.
44. The new stepmother prepared herself/himself to meet the family.
45. The favorite aunt censored herself/himself in front of her nephew.
46. The youthful grandmother planted a garden for herself/himself in the yard.
47. The shunned heiress cried herself/himself to sleep.
48. The grateful niece asked herself/himself how she could repay her aunt.
49. The old baroness covered herself/himself with jewels.
50. The busy landlady worked herself/himself into a frenzy.
51. The pompous chairwoman patted herself/himself on the back.
52. The tired milkmaid took it upon herself/himself to clean up.
53. The devout nun mumbled to herself/himself in church.

54. The head policewoman assigned herself/himself to the job.

55. The aspiring showgirl practiced the part by herself/himself before the show.

56. The daring sister flew herself/himself to Paris.

57. The jealous girlfriend told herself/himself not to worry.

58. The distinguished lady wrote an autobiography about herself/himself.

59. The inquisitive granddaughter found herself/himself in a bind.

60. The doubtful housewife sold herself/himself on the idea.

APPENDIX 2

"Match" and "Mismatch" Pronoun Sentences Presented during Experiment 2

1. The actress predicted that he/she would get the job.

2. The aunt heard that he/she had won the lottery.

3. The mother believed that he/she was sick again.

4. The princess decided that he/she would marry the American.

5. The queen knew that he/she would abdicate the throne.

6. The girl hoped that he/she would be asked to dance.

7. The nun believed that he/she would be asked to pray.

8. The stepmother denied that he/she was disliked by the children.

9. The milkmaid heard that he/she would be given a raise.

10. The waitress admitted that he/she had stolen the money.

11. The housewife decided that he/she would go shopping today.

12. The niece hoped that he/she would get to see her cousin.

13. The sister admitted that he/she had copied the homework.

14. The hostess hoped that he/she had invited the right people.

15. The showgirl admitted that he/she needed to work on the dance routine.

16. The uncle hoped that he/she had picked out a good wine.

17. The bachelor predicted that he/she would feel nervous during the wedding.

18. The chairman insisted that he/she would increase the company's performance.

19. The father decided that he/she would be lenient with his daughter.

20. The prince forgot that he/she had to attend the state function.

21. The king noticed that he/she had lost the support of the peasants.

22. The boy hoped that he/she could attend summer camp this year.

23. The priest heard that he/she was about to be appointed bishop.

24. The butler hinted that he/she would soon be replaced.

25. The stepfather decided that he/she would buy presents for the stepdaughter.

26. The waiter hoped that he/she would receive a big tip.

27. The chairman forgot that he/she had called a meeting.

28. The nephew hoped that he/she would receive the inheritance.

29. The brother decided that he/she would not wait for his sister any longer.

30. The policeman insisted that he/she would retire next year.

31. The mailman worried that he/she had forgotten to leave the letter.

32. The ballerina predicted that he/she would win the prize.

33. The wife agreed that he/she was happy with the new house.

34. The daughter decided that he/she would sneak out that night.

35. The lady regretted that he/she had to leave New York.

36. The girlfriend believed that he/she had finally been asked on a date.

37. The maid hoped that he/she would receive a big raise this year.

38. The duchess insisted that he/she was having an affair with the movie star.

39. The man insinuated that he/she would tell the authorities.

40. The grandmother realized that he/she was aging.

41. The bride realized that he/she looked very beautiful.

42. The stewardess hoped that he/she would make her flight on time.

43. The landlady decided that he/she would evict the noisy tenants.

44. The girlscout predicted that he/she would sell the most cookies.

45. The actress denied that he/she had seduced the director.

46. The aunt agreed that he/she had no talent for raising children.

47. The mother admitted that he/she would need a babysitter for the evening.

48. The husband agreed that he/she should help with the dishes.

49. The fireman knew that he/she had to escape in the next few minutes.

50. The son realized that he/she missed his parents.

51. The boyfriend admitted that he/she was in a jealous rage.

52. The man decided that he/she would accept the job at Microsoft.

53. The steward hoped that he/she could help the passenger.

54. The grandfather believed that he/she had voted for Truman.

55. The congressman hoped that he/she would win the next election.

56. The landlord denied that he/she had turned off the electricity.

57. The bachelor insisted that he/she would never get married.

58. The son promised that he/she would stay out of trouble.

59. The waiter worried that he/she would forget their order.

60. The congressman predicted that he/she would win the election easily.

*Semantically Anomalous Sentences and
Control Sentences Presented during
Experiment 2*

1. The horse raced down the track and won/bargained the race easily.

2. The boat sailed down the river and sank/coughed during the storm.

3. The car rolled down the hill and stopped/complained when it hit a tree.

4. The rocket shot into the sky and exploded/laughed in a fireball.

5. The team rested for several days and won/burned the championship.

6. The executive showed the report and frowned/floated after the meeting.

7. The applicant refused the job and complained/exploded to the government agency.

8. The man paid for the job and left/popped his work gloves behind.

9. The runner expected to win and lost/ducked the race instead.

10. The little girl read the sad story and cried/cracked about it for days.

11. The man told the truth and resigned/milked in disgrace.

12. The doctor rushed to the scene and helped/saluted to revive the victim.

13. The actress called from New Jersey and complained/sank to her agent.

14. The lawyer hurried through the proceedings and blundered/hit during his presentation.

15. The fielder tossed the ball and missed/unpacked his chance to tag out the runner.

16. The student lectured on honesty and lied/docked about his grades.

17. The corporation moved to New Jersey and grew/sneezed very rapidly.

18. The player passed the ball and scored/slept the touchdown.

19. The man phoned from Chicago and volunteered/collided for the job.

20. The boy told the joke and laughed/dropped at the punch line.

21. The spy watched for several days and escaped/barked over the fence.

22. The man mailed the letter and waited/steered for a reply.

23. The patient left in pain and died/floated the next day.

24. The dealer sold the car and complained/sank about the transmission.

25. The company sued for damages and lost/resigned the case.

26. The politician read the funny story and laughed/cried for several minutes.

27. The lady sent the flowers and blushed/popped when she thought about them.

28. The employee asked to retire and resigned/guessed from the company.

29. The barge floated across the lake and docked/laughed next to the warehouse.

30. The boxer pushed into the ring and fought/wrote three good rounds.

31. The car raced around the speedway and crashed/bargained into a wall.

32. The plane sailed through the air and landed/laughed on the runway.

33. The puppy rolled the ball and barked/applauded at it.

34. The arrow shot into the air and fell/re-signed into a haystack.

35. The patient rested for the night and recovered/collided quickly from the operation.

36. The professor asked a question and responded/docked with a thoughtful comment.

37. The manager refused the raise and quit/clawed his job.

38. The worker paid the tip and left/popped his wallet on the table.

39. The actress expected to star and failed/saluted in her role in the play.

40. The criminal stopped in New York and fled/sank into a nearby store.

41. The army battalion rushed to the front and retreated/bit after the battle.

42. The boy rushed to the hospital and collapsed/steered on the emergency room floor.

43. The flight attendant called to the cockpit and returned/milked to the cabin.

44. The old man hurried across the street and fell/flew down the stairwell.

45. The boy tossed the frisbee and ran/slept across the field.

46. The gambler dealt the cards and lost/broke the pot of money.

47. The family moved from the building and prospered/explored in their new house.

48. The guest passed the food and ate/hired some of the chocolate cake.

49. The attorney phoned from Chicago and relented/brushed about the deal.

50. The girl told a scary story and cried/wrote for several hours.

51. The lion watched for several hours and attacked/sank the sickly deer.

52. The man served the wine and drank/popped several glasses before dinner.

53. The lady left the money and disappeared/flushed without a trace.

54. The young couple sold the house and moved/landed into an apartment.

55. The lawyer sued for malpractice and won/bit the court case.

56. The man read the letter and responded/popped with an angry reply.

57. The banker sent the large check and celebrated/saluted the financial deal.

58. The logs floated down the river and sank/sneezed during the storm.

59. The baby showed the doll and smiled/bargained happily for several minutes.

60. The hockey player pushed off the ice and scored/lectured the winning goal.

61. The airplane raced through the sky and landed/laughed at the airport.

62. The ball rolled down the hill and broke/resigned into several pieces.

63. The bottle rolled down the hill and broke/resigned into several pieces.

64. The bullet shot out of the gun and missed/celebrated its target.

65. The horse rested in the stable and slept/saluted for several hours.

66. The woman asked for a divorce and cried/sank after receiving the papers.

67. The clerk refused a promotion and quit/floated in disgust.

68. The reporter paid for the information and smiled/barked in anticipation of the story.

69. The patient expected to recover and died/docked of complications.

70. The man stopped in the alley and fled/steered before the police arrived.

71. The soldier issued the order and saluted to/collided with his commanding officer.

72. The woman rushed down the hall and slipped/bit on a banana peel.

73. The broker called into the office and left/burned in a huff.

74. The dog hurried through the park and barked/yelled at a running squirrel.

75. The soldier tossed the grenade and ducked/danced into a foxhole.

76. The boy left in the car and whined/popped about his parents.

77. The actress moved to Los Angeles and unpacked/sank her bags.

78. The senator read the legislation and voted/flushed in favor of the measure.

79. The girl bought the nightgown and blushed/floated at the thought of wearing it.

80. The executive denied the request and quit/flew his job the next day.

81. The woman sent the bad news and fainted/broke into the arms of her husband.

82. The loan shark mailed the loot and disappeared/saluted for three years.

83. The passenger left on the boat and drowned/barked during the storm.

84. The old woman sold the farm and died/collided two months later.

85. The customer refused the seat and left/chomped a very small tip.

86. The actor read the poor review and cursed/brushed the newspaper critic.

87. The driver stopped at the light and waited/broke for the light to change.

88. The lawyer sent the good news and celebrated/steered with a bottle of wine.

89. The balloon floated in the air and popped/slept with a loud bang.

90. The lady pushed into the cab and complained/chopped about the rude service.

REFERENCES

- AISSIN, J. (1988). Extensions of brother-in-law agreement. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language*. Palo Alto: CSLI.
- BACH, E. (1974). *Syntactic theory*. New York: Holt, Reinhart, & Winston.
- BAKER, C. L. (1989). *English syntax*. Cambridge, MA: MIT Press.
- BARLOW, M. & FERGUSON, C. A. (1988). *Agreement in natural language*. Palo Alto: CSLI.
- BATES, E., & MACWHINNEY, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- BOCK, K., & CUTTING, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, **31**, 99–127.
- BOCK, K., & EBERHARD, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, **8**, 57–99.
- BOCK, K., & MILLER, C. A. (1991). Broken agreement. *Cognitive Psychology*, **23**, 45–93.
- CHAFE, W. (1970). *Meaning and the structure of language*. Chicago: University of Chicago Press.
- CHOMSKY, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- CHOMSKY, N. (1986). *Knowledge of language*. New York: Praeger.
- CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335–359.
- CORBETT, A., & CHANG, F. R. (1983). Pronoun disambiguation: Accessing potentials antecedents. *Memory and Cognition*, **11**, 283–294.
- CORBETT, G. (1991). *Gender*. Cambridge: Cambridge Univ. Press.
- CROFT, W. (1988). Agreement vs. case-marking and direct objects. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language*. Palo Alto: CSLI.
- DONCHIN, E. (1981). Surprise! . . . Surprise? *Psychophysiology*, **18**, 493–513.
- DONCHIN, E., & COLES, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, **11**, 343–427.
- DORON, E. (1988). On the complementarity of subject and subject–verb agreement. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language*. Palo Alto: CSLI.
- DUNCAN-JOHNSON, C. C., & DONCHIN, E. (1977). On quantifying surprise: The variation in event-related potentials with subjective probability. *Psychophysiology*, **14**, 456–467.
- FASSI FEHRI, A. (1988). Agreement in Arabic, binding and coherence. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language*. Palo Alto: CSLI.
- FRAZIER, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII*. Hillsdale, NJ: Erlbaum.
- FRAZIER, L., & RAYNER, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178–210.
- FRIEDMAN, D., SIMSON, R., RITTER, W., & RAPPIN, I. (1975). The late positive component (P300) and information processing in sentences. *Electroencephalography and Clinical Neurophysiology*, **38**, 225–262.
- GEISSER, S., & GREENHOUSE, S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**, 95–112.
- GERNSBACHER, M. A. (1991). Comprehending conceptual anaphors. *Language and Cognitive Processes*, **6**, 81–105.
- GIVON, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic*. New York: Academic Press.
- GIVON, T. (1979). *On understanding grammar*. New York: Academic Press.
- GLEASON, H. A. (1965). *Linguistics and English grammar*. New York: Holt, Reinhart, & Winston.
- GREENE, S. B., MCKOON, G., & RATCLIFF, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **2**, 266–283.
- HAGOORT, P., BROWN, C., & GROOTHUSEN, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, **8**, 439–484.

- HILLYARD, S. A., & PICTON, T. W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), *Handbook of physiology. Section I: Neurophysiology*. New York: American Physiological Society.
- HOLCOMB, P. J., & NEVILLE, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, **5**, 281–312.
- HUDDLESTON, (1984). *Introduction to the grammar of English*. Cambridge, UK: Cambridge Univ. Press.
- JASPER, H. H. (1958). Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, **10**, 371–375.
- JESPERSON, O. (1933). *Essentials of English grammar*. University, AL: University of Alabama Press.
- KUTAS, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, **8**, 533–572.
- KUTAS, M., & HILLYARD, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, **11**, 99–116.
- KUTAS, M., & HILLYARD, S. A. (1980b). Reading between the lines: Event-related brain potentials to grammatical errors and semantic anomalies. *Brain and Language*, **11**, 354–373.
- KUTAS, M., & HILLYARD, S. A. (1980c). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, **207**, 203–205.
- KUTAS, M., & HILLYARD, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, **11**, 539–550.
- KUTAS, M., & VAN PETTEN, C. (1988). Event-related potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology*. Greenwich, CT: JAI Press.
- LAPOINTE, S. G. (1988). Toward a unified theory of agreement. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language*. Palo Alto: CSLI.
- MACWHINNEY, B., BATES, E. S., & KLIÉGL, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, **23**, 127–150.
- MCKINNON, R., & OSTERHOUT, L. (1995). *Constraints on movement phenomena in sentence processing: Evidence from event-related potentials*. Manuscript submitted for publication.
- MUNTE, T. F., HEINZE, H., & MANGUN, G. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, **5**, 335–344.
- NEVILLE, H. J., KUTAS, M., CHESNEY, G., & SCHMIDT, A. L. (1985). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*, **25**, 75–92.
- NEVILLE, H. J., NICOL, J., BARSS, A., FORSTER, K. I., & GARRETT, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, **3**, 151–165.
- NICOL, J. (1988). *Coreference processing during sentence comprehension*. Unpublished doctoral dissertation, MIT.
- OAKHILL, J., GARNHAM, A., & VONK, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes*, **4**, S1 263–286.
- OSTERHOUT, L. (1990). *Event-related brain potentials elicited during sentence comprehension*. Unpublished doctoral dissertation, Tufts University.
- OSTERHOUT, L. (1994). Event-related brain potentials as tools for comprehending language comprehension. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.
- OSTERHOUT, L. (1995a). *On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences*. Manuscript submitted for publication.
- OSTERHOUT, L. (1995b). *Brain potentials reflect gender stereotypes*. Manuscript in preparation.
- OSTERHOUT, L., HOLCOMB, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, **31**, 785–806.
- OSTERHOUT, L., & HOLCOMB, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, **8**, 413–438.
- OSTERHOUT, L., & HOLCOMB, P. J. (1995). Event-related potentials and language comprehension. In M. G. H. Coles & M. D. Rugg (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition*. New York: Oxford Univ. Press.
- OSTERHOUT, L., HOLCOMB, P. J., & SWINNEY, D. A. (1994). Brain potentials to garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 786–803.
- OSTERHOUT, L., MCKINNON, R., BERSICK, M., & COREY, V. (1995). *On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family?* Manuscript submitted for publication.
- OSTERHOUT, L., NICOL, J., MCKINNON, R., NI, W., FODOR, F. D., & CRAIN, S. (1994). *An event-related brain potential analysis of the temporal course of syntactic and pragmatic anomaly detection*. Paper presented at the Seventh Annual CUNY Conference on Human Sentence Processing, March 17–19, New York.
- PALMER, F. (1978). *Grammar*. Harmondsworth: Penguin.
- POLLARD, C., & SAG, I. (1988). An information-based theory of agreement. In D. Brentari, G. Larson, & L. MacLeod (Eds.), *Parasession on agreement in grammatical theory*. Chicago: Chicago Linguistic Society.

- RADFORD, A. (1988). *Transformational grammar*. Cambridge, UK: Cambridge University Press.
- RAYNER, K., CARLSON, M., & FRAZIER, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, **22**, 358–374.
- REID, W. (1991). *Verb and noun number in English: A functional explanation*. London: Longman.
- SANFORD, A. J., & GARROD, S. C. (1989). What, when, and how? Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, **4**, S1 235–262.
- SELLS, P. (1985). *Lectures on contemporary syntactic theories*. Stanford: Center for the Study of Language and Information.
- VAN PETTEN, C., & KUTAS, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, **18**, 380–393.
- WIERZBECKA, A. (1985). Oats and wheat: The fallacy of arbitrariness. In J. Haiman (Ed.), *Iconicity in syntax*. Amsterdam: John Benjamins.
- (Received October 4, 1993)
(Revision received January 23, 1995)