

Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations

Seana Coulson, Jonathan W. King and Marta Kutas

Departments of Cognitive Science and of Neurosciences, University of California San Diego, CA, USA

Arguments about the existence of language-specific neural systems and particularly arguments about the independence of syntactic and semantic processing have recently focused on differences between the event-related potentials (ERPs) elicited by violations of syntactic structure (e.g. the P600) and those elicited by violations of semantic expectancy (e.g. the N400). However, the scalp distribution of the P600 component elicited by syntactic violations appears to resemble that elicited by rare categorical events (“odd-balls”) in non-linguistic contexts, frequently termed the P3b. The relationship between the P600 and the P3b was explored by manipulating the grammaticality of sentences read for comprehension, as well as two factors known to influence P3b amplitude: odd-ball probability and event saliency. Oddball probability was manipulated by varying the frequency of morphosyntactic violations within blocks of sentences, and event saliency was manipulated by using two types of morphosyntactic violations, one of which was more striking than the other. The results indicate that the amplitude of the P600, like the P3b, was sensitive to both the probability and saliency manipulations, and that the scalp distributions for the effect of probability and grammaticality are essentially similar. An unexpected, but not wholly surprising, finding was the elicitation of an anterior negativity between 300 and 500 msec post-word onset, which may index working memory operations involved in sentence processing.

Requests for reprints should be addressed to Marta Kutas, Department of Cognitive Science, University of California San Diego, La Jolla, CA 92093-0515, USA. E-mail: mkutas@ucsd.edu

The research reported in this paper was supported by NICHD Grant HD22614 and by NIA Grant AG08313 to M.K. In addition, S.C. was supported by funds from PHS Institutional Training Grant 5 T32 MH14268-19 to the Center for Human Information Processing (CHIP) at UCSD, and by a graduate fellowship from the San Diego McDonnell-Pew Center for Cognitive Neuroscience. J.W.K. was also supported in part by a McDonnell-Pew post-doctoral fellowship from the San Diego Center for Cognitive Neuroscience, and funds from NIH Grant T32 DC00041-01 to the Center for Research on Language (CRL) at UCSD. The authors would also like to thank Julie Boland, Robert Kluender and Tom Münte for their helpful suggestions on earlier drafts of this paper.

INTRODUCTION

The controversy surrounding whether syntactic processing is distinct from the rest of language processing has recently included evidence obtained from the brain's electrical response to words in sentence contexts. Osterhout, McKinnon, Bersick and Corey (1996) report that event-related brain potentials (ERPs) elicited by syntactic anomalies differ qualitatively from the brain response to semantic and pragmatic anomalies, and highlight the congruity of these results with a modularly organised language processor. We shall argue that the difference in the brain response to syntactic and semantic/pragmatic anomalies has been overstated. Furthermore, we present data which suggest that the ERP response most commonly linked to syntactic processing—a late positivity alternately dubbed the P600 (Osterhout, 1990; Osterhout & Holcomb, 1992) and the syntactic positive shift (Hagoort, Brown, & Groothusen, 1993)—is a member of the P300 family, a domain-general brain response elicited by rare and/or informative events (Pritchard, 1981).

The issue which has most polarised psycholinguists for the past two decades is the modularity of the language processor. This debate concerns a parallel set of issues in linguistics, psycholinguistics and neurolinguistics. In addition to the general question of whether knowledge about language is somehow special, and therefore separate from knowledge about other domains, linguists are concerned with whether different sorts of linguistic knowledge, such as phonology, morphology, syntax and semantics, have distinct and independent mental representations.

Psycholinguists have been more concerned with the status of the *processors* which compute linguistic representations. Specifically, the issue has been whether language processing is modular, separated from more general cognitive processes, and whether it can be broken down into the operations of various submodules, subserving operations on distinct linguistic representations. Empirical tests of the modularity thesis have focused primarily on determining when higher-level contextual information influences processing at lower levels. Proponents of modularity have sought to demonstrate the existence of low-level processors, such as parsers, which are strictly limited in their inputs and outputs and frequently construed as operating on categorical information. Proponents of interactive approaches have countered with demonstrations that processes believed to be modular are in fact influenced by the “wrong” types of information, or that their outputs are of a more graded character.

In its current incarnation, the debate is between those who favour strictly rule-based accounts of parsing and those who prefer constraint-based accounts. Advocates of rule-based accounts view the initial stages of sentence processing as the independent computation of phonological,

syntactic and lexical semantic representations via the application of rules (see Frazier, 1979, for a paradigmatic example). In contrast, advocates of constraint-based accounts portray sentence processing as the simultaneous application of different sorts of linguistic constraints needed to compute a message-level interpretation (e.g. Spivey-Knowlton & Tanenhaus, 1994).

Neurolinguists have imported these concerns into the brain, asking whether there are brain areas dedicated to language processing, and whether different sorts of linguistic processing can be assigned to different neural systems. Rule-based accounts predict that the qualitatively different processing principles for syntax and semantics arise from the existence of qualitatively different neural processors. Constraint-based accounts, on the other hand, suggest that sentence processing principles reflect general properties of the brain mechanisms subserving memory, perception and learning. For example, MacDonald, Pearlmutter and Seidenberg (1994) suggest that while there may be distinctly linguistic forms of representation, the processing principles that account for language comprehension involve constraint satisfaction procedures which apply across a broad range of cognitive domains. Moreover, differences in neural processing mechanisms dedicated to language are not wholly attributable to innate architectural constraints, but may instead reflect the interaction of informational properties of the input with general processing biases of the left hemisphere (Elman et al., 1996).

Evidence from Electrophysiological Measures

On the whole, hypotheses about the existence of qualitatively different syntactic and semantic processing mechanisms have been difficult to test with reaction time paradigms. Since both syntactic and semantic anomalies can result in an increase in processing time, measures such as reading time, lexical decision and word monitoring latency all yield similar results for the two classes of violations. These and other limitations have led many investigators to supplement existing behavioural techniques with scalp-recorded ERPs. Event-related brain potentials provide a continuous measure of cognitive—including language—processing in the absence of irrelevant task demands. Given certain assumptions about the relationship between representation, processing and neural activity, ERPs can in principle be used to discriminate between the operation of qualitatively different processing mechanisms (for assumptions, see Hagoort & Brown, 1994; Osterhout, 1994).

In the general case, ERPs are patterned voltage changes in the on-going electroencephalogram (EEG) that are time-locked to specific processing events, most commonly the onset of stimuli but also including the execution of motoric responses (Coles & Rugg, 1995; Hillyard & Picton, 1987). We

obtain ERPs by recording subjects' EEG and averaging the brain response to stimulus events; for example, the onset of the critical word in a large number of ungrammatical sentences. The logic behind averaging is to extract from the EEG only that information which is time-locked to the processing of the event. The result is a waveform containing a series of inflections that appear to the eye as "peaks" and "troughs", but which are usually all referred to as "peaks", since the voltage *polarity* (negative or positive) is explicitly taken into account. While such peaks are themselves often called "components", it should be understood that the notion of a component properly includes its "scalp distribution" (i.e. the pattern of relative amplitudes the component has across all recording sites). The "latency" of a peak is usually defined as the time point where the component reaches its largest amplitude. Thus a frequent convention for labelling peaks and components is to combine either the letter "N" or "P" depending on the peak polarity with the approximate peak latency in milliseconds; thus "N400" refers to a negative wave whose peak amplitude is at 400 msec after stimulus onset.

Cognitive neuroscientists have detailed many aspects of ERP components which correlate with cognitive operations involving memory, attention and even language (see Hillyard & Picton, 1987, for a review). The ERP approach seeks correlations between the dimensions of ERP components elicited by different stimuli and putatively relevant dimensions of the stimuli themselves. The observation of systematic modulations in the amplitude and/or latency of particular components can be used to test hypotheses about the underlying processing events. Although there are known limitations to using ERP data to localise neural generators in the brain, it is an excellent measure for determining precisely *when* the processing of two classes of stimuli begins to diverge.¹ Because brainwave measures are acquired with a high degree of temporal resolution (in the present case, every 4 msec), ERPs can potentially reveal the exact moment of divergence in the processing of particular categories of events. In any case, the detection at time *t* of a reliable difference in the ERP waveforms elicited by two categories of events suggests that processing of those categories differs at that instant, and began at least by time *t* (Coles, Gratton, & Fabiani, 1990).

Moreover, the multidimensional character of the ERP signal provides a number of ways to test hypotheses regarding qualitative differences in processing information at different representational levels (Hagoort & Brown, 1994). To the extent that processing at distinct representational

¹ Without a careful model of the neural generators responsible for a given set of ERP data (which must rely on auxiliary assumptions as well), one can only make very crude inferences about which brain regions are responsible for the generation of an ERP component. For example, the lateral asymmetry of an ERP component implies a lateral asymmetry in the generation process, but little else.

levels is subserved by anatomically distinct brain areas, we might expect different ERP components to index syntactic versus semantic processing. Representation-specific ERP components would thus support the independence of syntax and semantics. Similarly, language-specific ERP components would support the modular nature of the language processor. Naturally, identifying such an effect would require a relatively direct relationship between the linguistic processing of interest and the neural generator(s) of the ERP component. However, even in the absence of such a direct relationship, the experimenter can exploit properties of the electrical signal to help determine the relative independence of neural generators.

Event-related brain potentials recorded at the scalp reflect the summation of electrical activity from large numbers of neurons acting synchronously, primarily in the neocortex. Excitatory and inhibitory post-synaptic potentials create electrical fields which are volume-conducted to the scalp. Consequently, the amplitude of an ERP component can reflect the activity of more than one generator. However, because volume-conducted electrical fields summate where they intersect (Helmholz's Law of Superposition), independent generators will have additive effects on the amplitude of the scalp activity. Accordingly, the independence of representational levels can be tested by experimentally manipulating variables thought to engage different sorts of processing to see whether or not their effects are additive.

Another way to distinguish between representational levels is to look at the scalp distribution of a given ERP effect. Although effects visible at the scalp could be generated by a number of different configurations of neural generators, the reverse does not hold (see Dale, 1994, for additional discussion). Distinct scalp topographies strongly suggest some difference in the underlying neural generators. Consequently, we can manipulate variables thought to engage different sorts of processing to see whether the scalp distribution of the effects differs as a function of experimental condition.

ERP INDICES OF SEMANTIC AND SYNTACTIC PROCESSING

Given the assumption that qualitative differences in the ERP waveform reflect the operation of qualitatively different cognitive processes, previous researchers have sought to identify distinct components in the ERP which index specific linguistic processes (for reviews, see Hagoort & Brown, 1994; Osterhout, 1994). For example, the N400 component, a negative component peaking approximately 400 msec after word onset, is elicited by initial presentation of semantic anomalies (Kutas & Hillyard, 1980) and its amplitude is inversely related to cloze probability (see Bloom & Fischler,

1980, for a review of this concept) of a word in sentence context (Kutas, Lindamood, & Hillyard, 1984).

Although it was initially observed in response to violations of semantic expectancies, the N400 component is elicited by each word in a sentence (e.g. Kutas, Van Petten, & Besson, 1988). Semantic anomalies merely represent the endpoint along a continuum of semantic expectancy afforded by the message level representation. Thus, the reduction in N400 amplitude across the course of a sentence has been interpreted as reflecting a reduction in the difficulty of sentential integration due to the build-up of semantic constraints (Van Petten & Kutas, 1990).

Researchers have also identified two ERP components which seem to be sensitive to *syntactic* manipulations. Interestingly, the component with the shorter peak latency is a negativity which occurs in approximately the same time window as the N400, but is more anterior in its distribution. The second component, which has received more attention, is a broadly distributed positive shift beginning approximately 500 msec after the occurrence of a violation. This response has alternately been dubbed the P600 (Osterhout & Holcomb, 1992) and the syntactic positive shift (Hagoort et al., 1993). This component has been interpreted as indexing (directly or indirectly) the operation of a modular parser.

In the first study of this ilk, Osterhout and Holcomb (1992) examined subjects' ERPs to English sentences with two types of syntactic errors: deviation from preferred subcategorisation constraints and violations of phrase structure. Osterhout and Holcomb compared subjects' ERPs to (1) and (2).

1. The broker hoped to sell the stock.
2. The broker persuaded to sell the *stock.

Although sentence (2) does not become ungrammatical until the last word ("stock"), the word "to" is inconsistent with the preferred reading of "persuaded" as a past tense form that subcategorises for a noun phrase. Consequently, the word "to" presents the reader with a deviation from the verb's preferred subcategorisation frame, which may prompt syntactic reanalysis. Osterhout and Holcomb reported a slow, positive shift in the ERP, largest over the right anterior scalp, which was elicited by the apparent subcategorisation violation (at "to") in (2). They observed a similar positivity, albeit with a more posterior and laterally symmetric scalp distribution, to phrase structure violation at "was" in (4):

3. The broker persuaded to sell the stock was sent to jail.
4. The broker hoped to sell the stock *was sent to jail.

Although the positive-going component did not have a peak, it was labelled the P600 because its midpoint was about 600 msec after word onset.

Other investigators have since reported late positivities in the ERP which are sensitive to the well-formedness of linguistic stimuli. In particular, Hagoort et al. (1993) report similar effects elicited by phrase structure and subject–verb agreement violations in Dutch, and refer to the effect as the syntactic positive shift (SPS). Hagoort et al. argue that the sensitivity of the SPS to syntactic (but not semantic or pragmatic) violations points to a level of syntactic processing that is distinct from the semantic processes indexed by the N400. Neville et al. (1991) likewise report late positive effects in the ERP to subadjacency and phrase structure violations, in contrast to the N400 effect they observed to sentences which contained a *semantic* anomaly. Neville et al. argue that the different patterns of ERP activity elicited by syntactic and semantic violations clearly suggests a corresponding division between syntactic and semantic processing mechanisms.

Unfortunately, the data do not support a one-to-one mapping between N400 and semantics and P600 and syntax. For example, Hagoort et al. (1993) report a negative component peaking at 400 msec post-onset for their phrase structure violations. It is broadly distributed over the entire head, with a maximum over the posterior right scalp—the region where the N400 is usually the largest. Furthermore, while “to” in (2) (“The broker persuaded to sell the stock”) elicited a P600, “stock” elicited an N400 (Osterhout, 1990). This N400 may reflect the failure to assign a syntactic structure, concomitant semantic difficulty, or both. Moreover, in the following section, we point to data which argue against the interpretation of the P600/SPS as an unequivocal marker of syntactic processing.

How Syntactic is the Syntactic Positive Shift?

Thus far, results in the P600/SPS literature defy unified explanation in terms of syntactic theory. For example, Osterhout and Holcomb (1992) report a P600 in response to both phrase structure and subcategorisation violations. Although the ERPs to both violations displayed late positivity, the onsets and distributions of the two effects were different. Likewise, in Hagoort et al. (1993), the positivity elicited by the phrase structure violations began almost immediately after the onset of the stimulus, whereas the onset of the positivity elicited by the agreement condition was not until 500 msec post-stimulus, and there was no positive shift at all in their subcategorisation violation condition.

Furthermore, when we look across studies for a cohesive account of the late positivity elicited by syntactic violations, the data are even more resistant to unification (see Table 1). Phrase structure violations in Dutch elicit a positivity whose onset is almost immediate (Hagoort et al., 1993),

TABLE 1
ERP Indices of Syntactic Processing

| <i>Violation Type</i> | <i>References</i> | <i>Study Parameters</i> | | | <i>LPC Effects</i> | | | <i>Pre-LPC Negativities</i> |
|-----------------------|------------------------------|--------------------------|-----------------|-----------------------|---------------------|-----------------------------|-------------|-----------------------------|
| | | <i>Subject Task</i> | <i>Language</i> | <i>Latency (msec)</i> | <i>Distribution</i> | <i>Pre-LPC Negativities</i> | | |
| Agreement | Hagoort et al. (1993) | Reading | Dutch | >500 | Parietal | None | None | |
| Phrase structure | Neville et al. (1991) | Grammaticality judgement | English | 500–700 | Occipital | L. anterior | L. anterior | |
| | Osterhout and Holcomb (1992) | Grammaticality judgement | English | 500–700 | R. anterior | L. anterior | L. anterior | |
| | Hagoort et al. (1993) | Reading | Dutch | 100– | Broad | None | None | |
| Subcategorisation | Osterhout and Holcomb (1992) | Grammaticality judgement | English | 600–900 | Posterior | None | None | |
| | Hagoort et al. (1993) | Reading | Dutch | NA | NA | N400? | N400? | |
| | Rosler et al. (1993) | Lexical decision | German | NA | NA | L. anterior | L. anterior | |
| Specificity | Neville et al. (1991) | Grammaticality judgement | English | NA | NA | L. anterior | L. anterior | |
| Subjacency | Neville et al. (1991) | Grammaticality judgement | English | (700–) | Broad | None | None | |

NA, not applicable.

whereas the positivity to similar violations in English does not occur until 500 msec post-stimulus (Neville et al., 1991; Osterhout & Holcomb, 1992). Moreover, Neville et al. (1991) report a laterally symmetric positivity largest over occipital regions, whereas Osterhout and Holcomb (1992) report a positivity with a right anterior distribution for similar sorts of phrase structure violations. Finally, Osterhout and Holcomb (1992) report a positivity with a symmetric posterior distribution in response to violations of subcategorisation constraints in contrast to Hagoort and co-workers' (1993) null result for subcategorisation violations in Dutch.

Whereas differences in the latency of an ERP component do not necessarily imply qualitative differences in the response, differences in the scalp distribution imply differences in the underlying neural generators.² Given the variety in the scalp distribution of the ERP responses to syntactic violations within and across studies, it is far from clear how we can systematically map the disparate ERP responses onto either general syntactic principles or properties of the specific languages involved. While some of these discrepancies can be explained away by recourse to models that predict differences in the neural implementation of different languages (e.g. Bates & MacWhinney's, 1989, Competition Model) not all of the observed differences in brain response to syntactic violations can be reduced to differences between English, Dutch and German. Especially disconcerting is the observation (see Table 1) of a preceding negativity whose intermittent presence has yet to be cogently explained (cf. Kutas & Kluender, 1994; Rösler, Pütz, Friederici, & Hahne, 1993).

Results from the literature are listed in Table 1 according to the sort of syntactic violation each study addressed and the latency and distribution of the positivity found in the ERP. The right-hand column shows whether or not the observed positivity was preceded by a negative going component.

The variety of ERP components elicited in response to syntactic manipulations indicates that there is no *unique* index of syntactic processing. The most consistent result is for syntactic anomalies to elicit a late positivity in the ERP with a laterally symmetric, roughly posterior distribution. However, there is no particular reason to accept the so-called syntactic positive shift as syntax-specific.

The very existence of separate components in the waveform which index functionally distinct levels of processing might be seen by some as implicit support for the modularity thesis. If syntactic and semantic processing are subserved by independent modules, it would be quite natural to expect different aspects of processing to be manifested in qualitatively different

² Note, however, that ERPs reflect the contribution of many generators, and that latency shifts in the activity of one or more of these could generate differences in scalp distribution where there is actually no difference in the identity of generators involved.

ERP effects. Once identified, an ERP index of syntactic processing could be examined in various contexts to reveal the extent to which it is modulated by semantic processing. An ERP index of syntactic processing would also be especially useful for evaluating claims about the time-course of processing.

While both Osterhout and Holcomb (1993) and Hagoort et al. (1993) carefully note the difficulty involved in making transparent inferences from the existence of an ERP component to claims about the modularity of the syntactic processing system or its neural underpinnings, a cursory reading of abstracts could mislead even the most sensible reader to think that these findings argue for a modular parser. The current study is intended to clarify the relationship between the P600/SPS and the P3 family of ERP components, and to evaluate the utility of the P600 as a dependent measure of syntactic processing.

P600/SPS and the P3b

Cognitive neuroscientists familiar with the attention and decision-making literatures will see similarities between the P600/SPS and one of the earliest known ERP components, the P300. We suggest that the late positivity time-locked to syntactic irregularity is actually a member of the P300 family, a host of positive components with varying onsets and peak latencies. The most heavily investigated component in this family is the P3b, a positive going component with centro-parietal maximum. The P3b is known to reflect the resolution of prior uncertainty and the task-relevant surprise value of the stimulus. For example, in the auditory odd-ball paradigm, in which the subject is directed to attend to a series of long tones periodically interspersed with short tones (or vice versa), the rarer stimuli elicit a robust P3b. The amplitude of this response is proportional to the rarity of the target stimulus and its latency varies with the difficulty of the categorisation task (Picton, 1992).

The discovery of the P3b component was a watershed in the field of psychophysiology because this component is sensitive to cognitive aspects of processing, rather than being a purely sensory reaction to the stimuli themselves. While the scalp distribution of the P3b does vary somewhat as a function of stimulus modality (Johnson, 1989), P3b amplitude variations are associated with variations in *subjective* aspects of the stimuli, such as their task relevance, salience and probability (for reviews, see Johnson, 1986; Pritchard, 1981). Thus the P3b does not primarily reflect the physical dimensions of eliciting stimuli, but the perception of those dimensions. One proposal as to the functional role of the cognitive process or processes associated with the P3b is the “context updating” model. On this model, the P3b indexes processes involved in updating an individual’s mental model of

the environment (Donchin, 1979; Donchin & Coles, 1988; but see Verleger, 1988, for an alternative account).

The P600/SPS, an ERP response previously associated with purely syntactic processing, might in fact be a P3b elicited by the relative rarity of words which make sentences ungrammatical. The identification of the P600/SPS with the P3b is suggested by the similar scalp distributions of the two components. While our previous discussion demonstrates that the response to syntactic anomaly is hardly monolithic, the frequent observation of a centro-parietally distributed P600/SPS is consistent with the scalp distribution of the P3b. The variety in scalp distributions of positivities elicited by syntactic anomalies might also reflect overlap of other ERP effects which vary with other processes triggered by different sorts of violations. Furthermore, the change in P600 distribution observed by Osterhout and Holcomb (1993) for visual versus auditory presentation is similar to that normally observed for the visual versus the auditory P3b (Johnson, 1989). Specifically, the P3b and the P600 elicited by visual stimuli tend to be larger over the frontal and central scalp than the ERP to auditory stimuli.

To test the relationship between the P600/SPS and the P3b, we used a paradigm designed to elicit both components. To elicit the P600/SPS, we contrasted sentences containing morphosyntactic violations with sentences containing no such violations. And because one of the chief characteristics of the P3b is an amplitude increase with decreasing stimulus probability, we varied the probability of occurrence of syntactic violations between blocks of sentences. Although the stimulus manipulation was linguistic in character, our probability manipulation used the odd-ball paradigm in much the same way as in the attention literature, where researchers manipulate the relative proportion of categorisable stimuli (e.g. the proportion of long vs short tones) to render one category more surprising. By varying the probability of ungrammatical sentences, we hoped to induce participants to expect grammatical sentences when most sentences were grammatical, and ungrammatical sentences when most sentences were ungrammatical. Since we also expected that *saliency* of the violations would have an effect on P3b amplitude, we used two different kinds of morphosyntactic violations (pronoun case and subject-verb agreement) that differed noticeably in saliency.

In our paradigm, the effect of grammaticality on the ERP is attributed to the neural generators of the P600/SPS, while the effect of probability is attributed to neural generators of the P3b. Moreover, because the physical properties of the electrical signal are well-understood, the ERP methodology affords two methods for comparing the brain response to grammaticality and to probability (see Osterhout et al., 1996, for similar reasoning). First, if the same neural generators underlie the P600/SPS and

the P3b, the grammaticality and probability effects should have the same scalp distribution. Second, because independent neural generators have additive effects on the amplitude of the ERP, we can infer the relationship between underlying neural generators from the additivity or non-additivity of grammaticality and probability effects.

If the P600/SPS and P3b are generated by different neural generators, which just happen to look similar at the scalp, the effects of grammaticality and probability should be additive. However, if the P600/SPS and the P3b are generated by essentially the same neural substrate, the grammaticality and probability effects should interact. On this latter hypothesis, the P600/SPS results from the violation of a default expectation of grammaticality and the modification of the model which supports those expectations. Consequently, one would expect to see a small probability effect on grammatical stimuli and a much larger probability effect on ungrammatical stimuli; namely, stimuli which tend to elicit P3b even when no special measures are taken to make them salient. Moreover, because the P3b is modulated by the salience of experimental stimuli, the identity of the two components predicts that more salient pronoun case violations should elicit greater amplitude P600/SPS than less noticeable subject-verb agreement errors.

METHODS

Event-related Brain Potentials

Event-related brain potentials were recorded from 26 geodesically arranged electrode sites on a standard electrocap, the right mastoid, and a non-cephalic derivation consisting of an average of the signal from one electrode at the sterno-clavicular junction, and one on top of the seventh cervical vertebra. Electrodes were placed at the outer canthi and under both eyes to detect eye movements and blinks. All electrodes were referenced to the left mastoid.

The electroencephalogram (EEG) was analog-filtered between 0.01 and 100 Hz and digitised at 250 Hz. Epochs with blocked channels, eye movements, or blinks were rejected off-line before averaging. This resulted in the rejection of 16.3% of all trials.

Participants

There were 16 participants (10 males, 6 females), all of whom were right-handed, monolingual English speakers, aged 18–30 years. All participants had normal or corrected-to-normal vision; none had any history of reading problems or neurological disorders.

Materials

The EEG was recorded as the participants read 200 different English sentences for comprehension. The ungrammatical sentences included one of two types of morphosyntactic violations: violations of subject–verb number agreement on the verb, and violations of overt case marking on pronouns. The words in each ungrammatical sentence were identical to those in its grammatical control, except for the single word which contained the morphosyntactic violation. Consequently, ERPs were time-locked to the presentation of the word which made the sentence ungrammatical (or to its grammatical counterpart). Examples of each type of violation and their respective controls are displayed below.

Pronoun case

- 5a. The plane took *we to paradise and back.
- 5b. The plane took us to paradise and back.
- 6a. Ray fell down and skinned *he knee.
- 6b. Ray fell down and skinned his knee.

Verb agreement

- 7a. Every Monday he *mow the lawn.
- 7b. Every Monday he mows the lawn.
- 8a. They *suns themselves on the beach.
- 8b. They sun themselves on the beach.

The 200 sentences were divided into two blocks of 100 sentences each. However, the two blocks varied in the proportion of grammatical to ungrammatical sentences. In the grammaticality probable block (ungrammaticality improbable), 20 of the sentences were ungrammatical and 80 were grammatical. In the grammaticality improbable block (ungrammaticality probable), 80 of the sentences were ungrammatical and 20 were grammatical.

Each block contained 100 sentences, 50 of which were pronoun case sentence frames and 50 verb agreement sentence frames. The number of ungrammatical sentences for each of the two sentence types (pronoun case vs verb agreement) depended on whether the block was a grammaticality probable (ungrammaticality improbable) block or a grammaticality improbable (ungrammaticality probable) block. In grammaticality probable (ungrammaticality improbable) blocks, there were 10 ungrammatical pronoun case sentences and 10 ungrammatical verb agreement sentences, 40 grammatical pronoun case sentences and 40 grammatical verb agreement sentences. In the grammaticality improbable (ungrammaticality probable) block, there were 40 ungrammatical pronoun case sentences, 40

ungrammatical verb agreement sentences, 10 grammatical pronoun case sentences and 10 grammatical verb agreement sentences.

The stimuli were divided into four lists which were varied across participants. The lists were designed so that order of blocks was counterbalanced across subjects and so that no subject saw the same sentence frame twice. However, when collapsed across subjects, comparison of ERPs with grammatical versus ungrammatical sentences represents the response to sentences which were identical up to (and after) the critical word.

Procedure

The sentences were presented visually one word at a time in the centre of the screen. Individual words were presented for 200 msec with an inter-stimulus interval (ISI) of 300 msec. After each sentence, there were 3 sec of blank screen during which the subject was allowed to blink and move. Before each sentence, a fixation symbol appeared for an interval which varied between 500 and 1200 msec. The task was to read 200 English sentences and answer 20 randomly interspersed comprehension questions.

RESULTS

Both the manipulation of stimulus grammaticality and the block probability led to ERP effects consistent with those reported in the literature. The ERPs to ungrammatical stimuli displayed slightly enhanced negativity in the 300–500 msec time window after the onset of the critical word, and increased positivity thereafter (see Fig. 1). The ERPs to improbable stimuli also displayed slightly enhanced negativity in the 300–500 msec time window post-word onset, and increased positivity from 500 to 800 msec (see Fig. 2).

Late Positive Effects: Grammaticality and Violation Type

The brain response to ungrammatical stimuli displayed the characteristic P600/SPS between 500 and 800 msec after word onset, peaking at 700 msec, and was largest over the central-parietal scalp. While researchers in psycholinguistics have traditionally employed the mean amplitude of the waveform to measure the P600/SPS, researchers in psychophysiology have traditionally employed baseline-to-peak amplitude to measure the P3b. Neither measure is ideal. Peak amplitude, defined as the maximum value at each electrode within some predefined time window, can lead to two problems. First, because it uses only a single data point at each electrode, peak amplitude is highly sensitive to noise. Second, because the

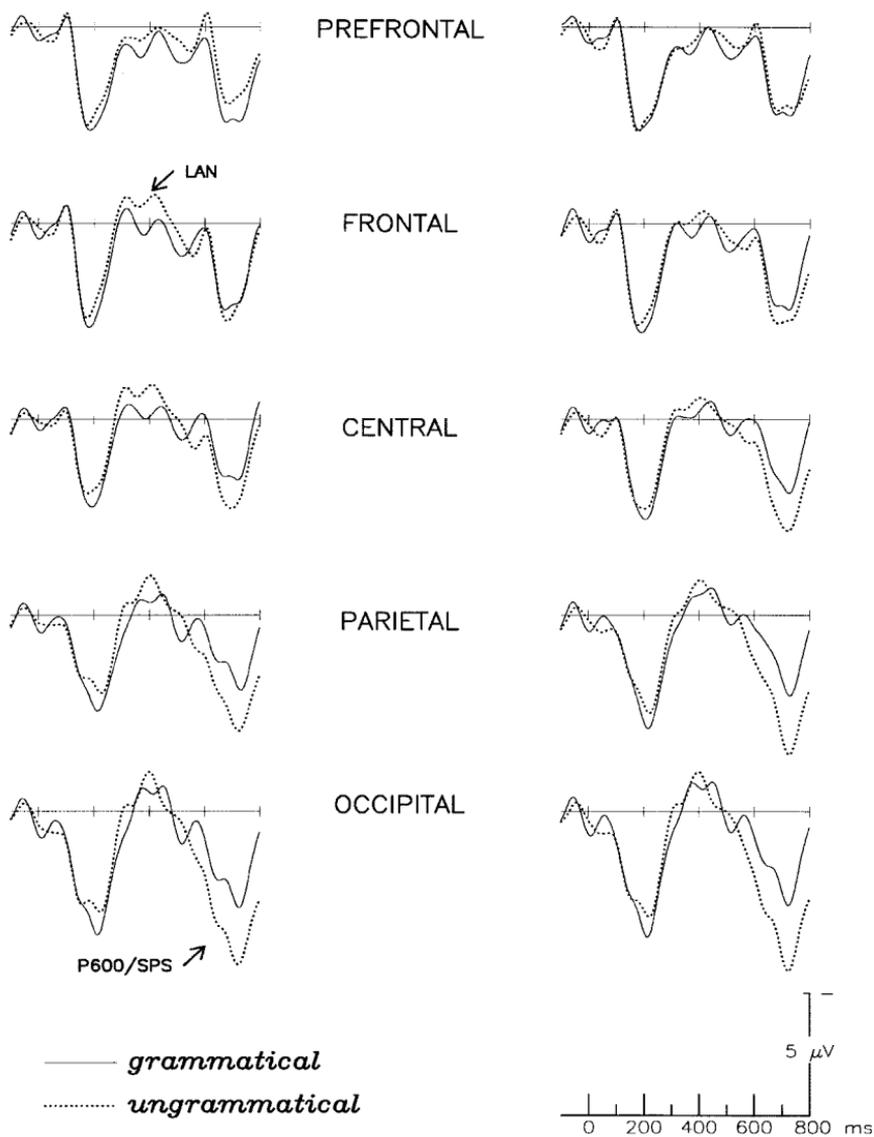


FIG. 1. Grammaticality effect. Grand average ERPs ($n = 16$) to critical words in grammatical and ungrammatical sentences. Compared to grammatical controls, ERPs to ungrammatical stimuli displayed slightly enhanced negativity in the 300–500 msec time window and increased positivity thereafter. Electrode sites pictured were recorded over the left and right prefrontal, frontal, central, parietal and occipital scalp sites. Unless otherwise noted, electrodes recorded from the left-hand side of the scalp are shown on the left-hand side of the figure. Negative voltage is plotted up in this and all subsequent figures.

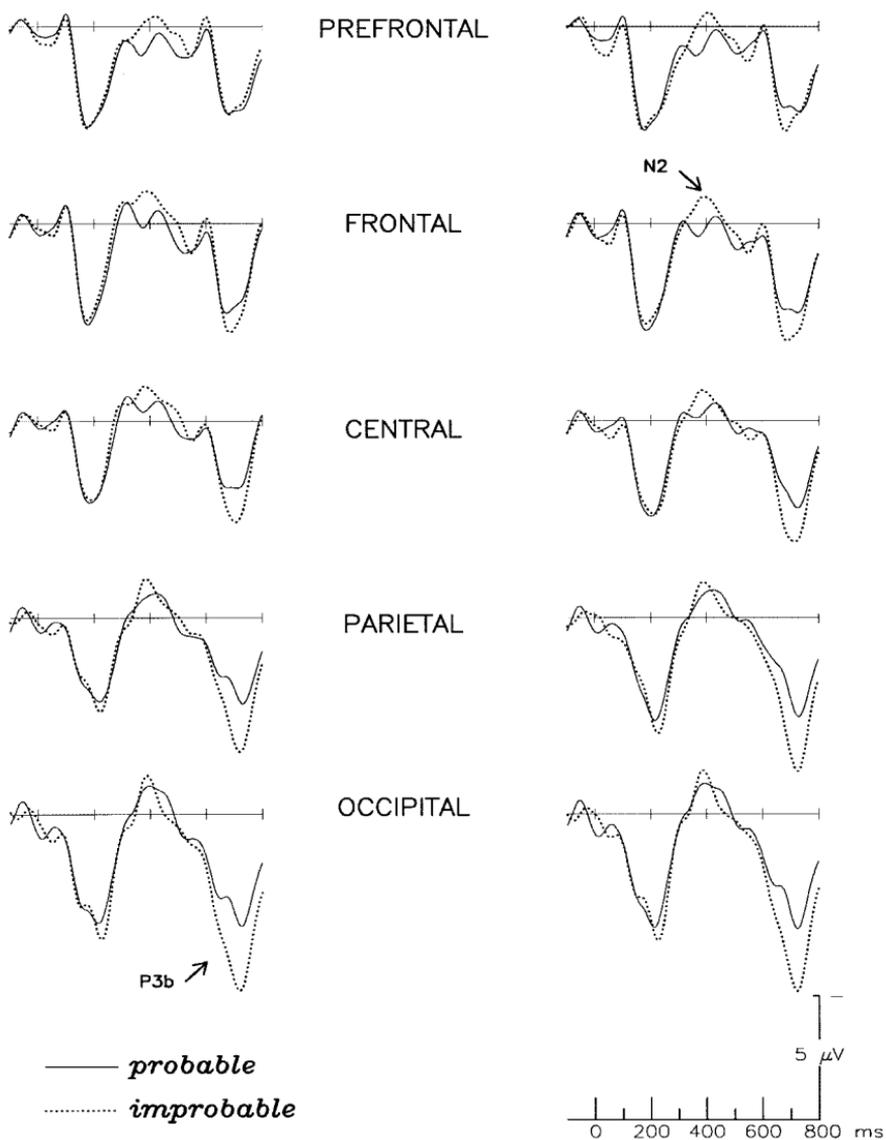


FIG. 2. Probability effect. Grand average ERPs ($n = 16$) recorded from the left and right prefrontal, frontal, central, parietal and occipital scalp sites. Compared with probable stimuli, ERPs to improbable stimuli displayed slightly enhanced negativity 300–500 msec, and enhanced positivity 500–800 msec, after word onset. Probable stimuli were grammatical in grammatical probable blocks, and ungrammatical in grammatical improbable blocks.

measurement at each electrode is independent, it is often the case that peak amplitudes at different electrode sites occur at different latencies—even for a single individual. This makes it all the more unlikely that the peak amplitude measure reflects the size of the same component (i.e. underlying generator) across individuals and electrode sites. As psychophysicologists have come to understand scalp distribution as an essential aspect of ERP components, there has been a concomitant realisation that the appropriate way to measure the peak amplitude of a component is to measure all channels at the same latencies.

In contrast to peak amplitude, mean amplitude is measured in the same time window at all electrode sites, by choosing one interval which will be appropriate for all participants. Unfortunately, broadening the time window enough to encompass inter-participant variability can increase the probability that the chosen interval encompasses more than one component. This is especially true in the case of the P3b, whose latency is known to be modulated by the complexity of stimulus processing, and often occurs in a complex with a preceding N2 and a following slow wave component (see Coles et al., 1990).

In an effort to index the true amplitude of the P3b component correctly, we developed a new measure which incorporates the features of both mean and peak amplitude measures, while attenuating their pitfalls. The hybrid measure which we developed uses a relatively narrow (100 msec) window around the peak of the positivity elicited at a midline parietal site where the P3b is known to be largest. Our mean around the peak measure increases the chance of measuring the waveform in the temporal region of interest, while simultaneously reducing noise.

To assess the reliability of the difference between ERPs to grammatical and ungrammatical events, we first measured the peak latency of the late positive peak at the mid-line parietal site to determine a 100 msec time window around this peak for each participant. Subsequently, we measured the mean amplitude of the waveform in this time window for every site.³ Reliability of effects was tested with an overall ANOVA with repeated measures, including violation type (pronoun case *vs* verb agreement),

³ Readers who are uneasy at the prospect of changing the time window around the peak on a subject by subject basis, will be pleased to know that the results are essentially similar if a fixed window from 650 to 750 msec after word onset is chosen (a 100 msec window around the mean peak latency of the late positivity). A repeated ANOVA analysis of the mean amplitude of the waveform in this time window with factors violation type, grammaticality, probability and electrode site ($n = 26$) yielded reliable effects of violation type, grammaticality and probability, interactions between violation type and grammaticality, violation type and electrode site, grammaticality and electrode site, probability and electrode site, and (most importantly) a three-way interaction between probability, grammaticality and electrode site (all at the 0.05 level or better, after the application of the Huynh-Feldt correction).

grammaticality (grammatical vs ungrammatical), probability (probable vs improbable), hemisphere (left vs right), laterality (lateral vs medial) and anteriority (4 levels) as factors. The Huynh–Feldt (1976) correction was applied where the use of repeated measures could lead to violations of sphericity in the ANOVA; we report nominal degrees of freedom but corrected P -values in the results below.

Analyses revealed reliable interactions between violation type and various scalp distribution factors [violation type \times laterality: $F(1,15) = 9.67$, $P < 0.01$; violation type \times anteriority: $F(1,15) = 6.93$, $P < 0.05$], reflecting a difference in the amplitude and distribution of the ERPs elicited by pronoun case and verb agreement stimuli. The ERPs elicited by pronoun case stimuli were more negative over anterior regions, and more positive over posterior regions, than those elicited by the verb agreement stimuli. Furthermore, ERPs elicited by critical words in pronoun case sentences were larger medially than laterally, while those elicited by verb agreement sentences were more evenly distributed over the scalp. These differences in the ERPs are not surprising, given that in our stimulus set the case errors all involved pronouns, while the agreement errors all involved verbs. Pronouns, members of the closed class, are shorter and much more frequent than verbs, members of the open class. The variables of length and frequency are known to affect the ERP elicited by words in both sentence contexts (Van Petten & Kutas, 1990) and in word lists (Rugg, 1990).

There was also a reliable effect of grammaticality, reflecting the larger positive response to ungrammatical stimuli [grammaticality: $F(1,15) = 7.03$, $P < 0.05$]. The late positivity to ungrammatical stimuli was largest at medial sites [grammaticality \times laterality: $F(1,15) = 20.98$, $P < 0.05$], and differed along the anterior–posterior dimension [grammaticality \times anteriority: $F(3,45) = 24.43$, $P < 0.001$]. Between 500 and 800 msec after word onset, ERPs elicited by grammatical stimuli were most positive in anterior regions, while those elicited by ungrammatical stimuli were most positive over the posterior scalp (see Fig. 3).

Grammaticality interacted with violation type [violation type \times grammaticality: $F(1,15) = 12.00$, $P < 0.01$], reflecting a larger, more broadly distributed effect of grammaticality in the pronoun case stimuli than in the verb agreement stimuli [violation type \times grammaticality \times laterality \times anteriority: $F(3,45) = 3.50$, $P < 0.05$] (see Fig. 4). *Post-hoc* tests suggested a grammaticality effect in verb agreement sentences, where ungrammatical stimuli elicited more positive ERPs in centro- and occipito-parietal sites, but less positive ERPs over the most anterior scalp regions [grammaticality \times anteriority: $F(3,45) = 6.44$, $P < 0.05$]. Ungrammatical pronoun case stimuli also elicited more positive ERPs than did their grammatical controls [grammaticality: $F(1,15) = 15.49$, $P < 0.01$]. The late positive response elicited by ungrammatical pronouns was largest over the posterior scalp

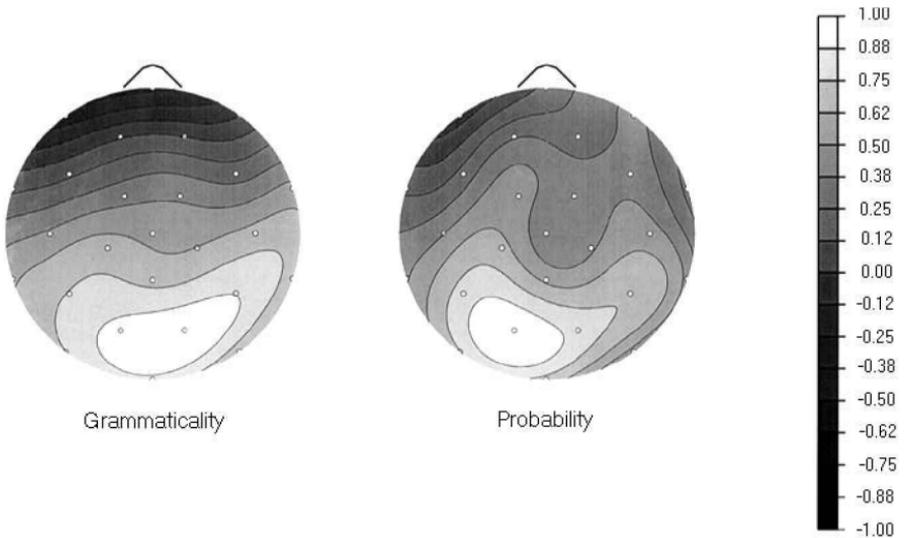


FIG. 3. Scalp distribution of grammaticality and probability effects. Topographic map of the mean amplitude of the grammaticality (ungrammatical minus grammatical) and probability (improbable minus probable) difference waves 700–720 msec after the onset of the critical word. The scalp distribution of the late positive effects of grammaticality and probability are very similar over the posterior scalp. Over the anterior regions, however, there are subtle differences in the scalp distribution of the two effects.

regions [grammaticality \times anteriority: $F(3,45) = 18.19$, $P < 0.001$], and larger over right than left anterior sites [grammaticality \times hemisphere \times anteriority: $F(3,45) = 3.79$, $P < 0.005$].

Late Positive Effects: Probability

Recall that probable stimuli were in grammatical sentences in grammatical probable blocks and in ungrammatical sentences in grammatical improbable blocks. The ERPs to improbable stimuli were more positive than the ERPs to probable stimuli [probability: $F(1,15) = 6.95$, $P < 0.05$], especially at lateral sites [probability \times laterality: $F(1,15) = 16.18$, $P < 0.01$]. The enhanced late positivity to improbable stimuli is characteristic of the P3b, and reveals participants' sensitivity to our manipulation of the local probability of grammatical and ungrammatical sentences. *Post-hoc* tests revealed that the probability effect for pronoun case stimuli was negligible over the most anterior and posterior sites, but prominent over the centro-parietal scalp [probability \times anteriority: $F(3,45) = 5.62$, $P < 0.01$].

Moreover, the probability manipulation differentially affected the ERP to grammatical and ungrammatical stimuli (see Fig. 5). While the positivity

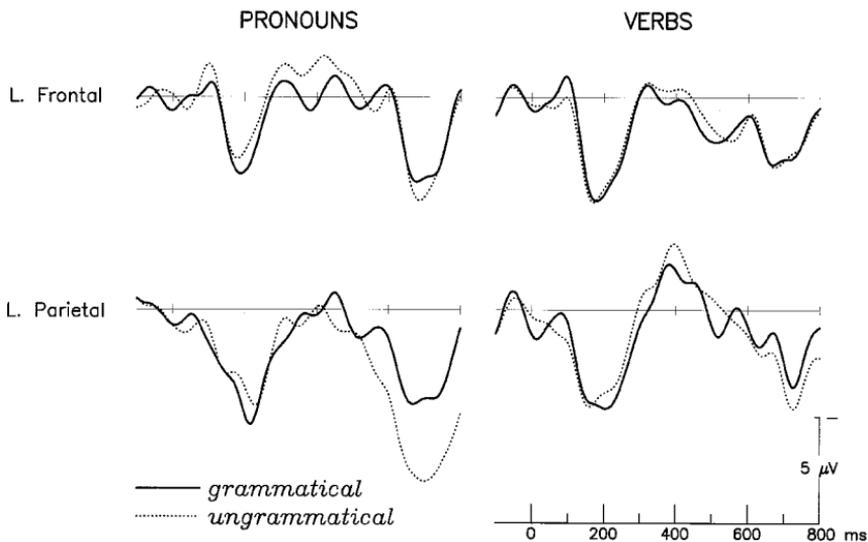


FIG. 4. Grammaticality effect in pronouns and verbs. Grand average ERPs elicited by critical pronouns (left) and verbs (right) at the left frontal and parietal electrode sites. The grammaticality manipulation had a much greater effect on the ERPs to the more salient pronoun case violations than to verb agreement violations. The positivity between 500 and 800 msec was much larger and more broadly distributed for ungrammatical pronouns than for ungrammatical verbs. The negativity between 300 and 500 msec post-onset was larger at left anterior sites in pronouns, and more posteriorly distributed in ERPs elicited by verbs.

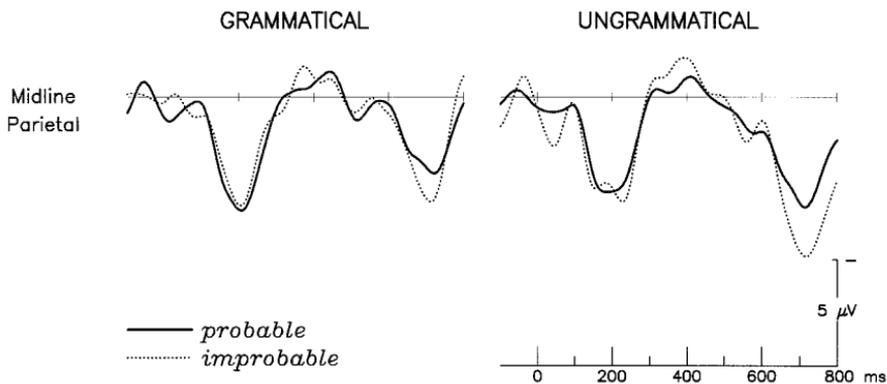


FIG. 5. Grammaticality \times probability interaction. Grand average ERPs ($n = 16$) elicited by the critical word, recorded from the mid-line parietal electrode site. In both grammatical and ungrammatical sentences, the probability manipulation enhanced the positivity elicited 500–800 msec after the onset of the critical word. However, the probability effect was more pronounced for words in ungrammatical than in grammatical sentences.

elicited by improbable grammatical stimuli was only slightly larger than that elicited by probable grammatical stimuli, the positivity elicited by improbable ungrammatical stimuli was much larger than that for probable ungrammatical stimuli, especially at centro-parietal sites where the P3b is usually largest [grammaticality \times probability \times hemisphere \times laterality: $F(1,15) = 5.28, P < 0.05$]. Among the grammatical stimuli, there was a small probability effect visible at all sites; among ungrammatical stimuli, the probability effect was not evident at the most anterior sites, but quite marked over the central and posterior scalp [grammaticality \times probability \times anteriority: $F(3,45) = 5.43, P < 0.05$].

Furthermore, the interaction between grammaticality and probability also differed as a function of violation type [violation type \times grammaticality \times probability: $F(1,15) = 4.76, P < 0.05$] (see Fig. 6). Among the grammatical verb agreement stimuli, the probability effect was frontally distributed, while the probability effect for ungrammatical items was more posterior [grammaticality \times probability \times anteriority: $F(3,45) = 5.70, P < 0.05$]. Among the grammatical pronoun case stimuli, the probability effect was bilaterally symmetric; the probability effect among the ungrammatical pronoun case stimuli was both larger, and more evident over the left hemisphere sites [grammaticality \times probability \times hemisphere \times laterality: $F(1,15) = 5.32, P < 0.05$].

One might argue that the observed interaction between grammaticality, probability and variables which reflect electrode site demonstrates that the scalp distributions of the grammaticality and probability effects are different, and therefore that different configurations of sources underlie the two effects (see, e.g. Osterhout et al., 1996). However, the inference of distinct neural generators from significant condition \times electrode site interactions on ANOVA tests is not straightforward. Because simple changes in source strength (i.e. main effects) have a non-linear effect on ERP amplitudes across the scalp, significant electrode site \times condition interactions can result from simple changes in source strength as well as from genuinely different configurations of sources (McCarthy & Wood, 1985).

To test for real differences in the scalp distribution of the two effects, values from our mean around the peak measure were normalised according to the method prescribed by McCarthy and Wood (1985). These values were then subjected to repeated measures ANOVA with factors violation type (pronoun case *vs* verb agreement), grammaticality (grammatical *vs* ungrammatical), probability (improbable *vs* probable) and electrode site (26 locations). This analysis yielded reliable interactions between violation type and electrode site [$F(25,375) = 3.08, P < 0.01$] and between grammaticality and electrode site [$F(25,375) = 6.11, P < 0.001$], suggesting different scalp distributions for pronoun case and verb agreement stimuli, and for

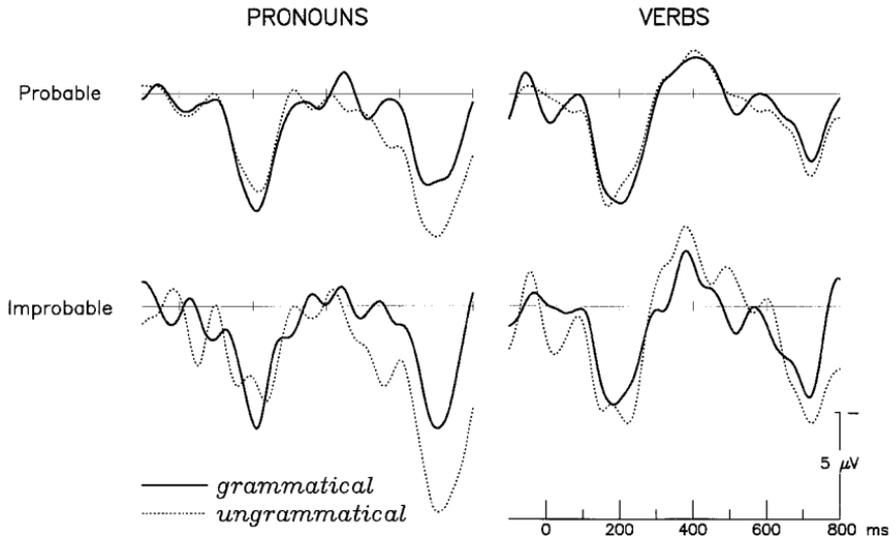


FIG. 6. Grammaticality \times probability interaction in verbs and pronouns. Grand average ERPs elicited by critical pronouns and verbs at the mid-line parietal site. Ungrammatical stimuli always elicited a larger positivity than grammatical stimuli 500–800 msec after word onset. However, the size of the grammaticality effect differed as a function of violation type and probability. The more salient pronoun violations elicited larger positivities, especially when they were improbable. The less noticeable verb agreement violations elicited a smaller grammaticality effect, which was only slightly enhanced in the improbable condition.

grammatical and ungrammatical stimuli. However, the interaction between grammaticality, probability and electrode site did not reach significance [$F(25,375) = 1.33, P = 0.24$]. Moreover, *post-hoc* tests of individual violation types revealed a similar pattern for the pronoun case stimuli and the verb agreement stimuli: the scalp distribution of the ERP differed as a function of the stimulus grammaticality, but the distributions of the probability and grammaticality effects did not differ reliably from each other [pronouns: grammaticality \times electrode site, $F(25,375) = 3.94, P < 0.01$; grammaticality \times probability \times electrode site, $F(25,375) = 0.87$; verbs: grammaticality \times electrode site, $F(25,375) = 1.44, P = 0.20$].

Early Negative Effects

In addition to replicating the effects of grammaticality on the P600/SPS noted by others, we can also report a replication of those studies that found an anterior negativity elicited by ungrammatical stimuli. As suggested by Fig. 1, this effect was largest between 300 and 500 msec after stimulus onset, which is also a convenient time window in which to measure the mean

amplitude of this waveform. These mean amplitudes were submitted to an ANOVA identical in all respects to that used in our analysis of the P600/SPS data.

The scalp distribution of the waveform in this earlier time window differed reliably as a function of violation type. The negativity elicited by pronouns was largest at left anterior sites, while that elicited by the verbs was centro-parietal and slightly larger over the right hemisphere [violation type \times hemisphere: $F(1,15) = 7.58$, $P < 0.05$; violation type \times anteriority: $F(3,45) = 12.96$, $P < 0.01$; violation type \times laterality: $F(1,15) = 18.86$, $P < 0.001$].

Furthermore, violation type interacted with grammaticality in a way which suggests differences in the scalp distribution of the grammaticality effect in the pronoun case stimuli and the verb agreement stimuli [violation type \times grammaticality \times laterality \times anteriority: $F(3,45) = 2.96$, $P < 0.05$] (see Fig. 7). *Post-hoc* tests revealed a main effect of grammaticality in the verb agreement stimuli, because the ungrammatical verb agreement stimuli elicited larger amplitude negativities [grammaticality: $F(1,15) = 16.41$, $P < 0.01$]. Ungrammatical pronoun case stimuli elicited enhanced frontal



FIG. 7. Grammaticality effect in pronouns and verbs: early negativity. Topographic map of mean amplitude of grammaticality difference waves (ungrammatical minus grammatical) elicited by critical pronouns (left) and verbs (right), 400–420 msec after word onset. The values have been normalised to correct for amplitude differences and interpolated to estimate voltage in regions between the 26 recording sites. The scalp distribution of the early negativity was more anterior for pronouns and more posterior for verbs.

negativity relative to grammatical controls, especially over left lateral scalp sites [grammaticality \times laterality \times anteriority: $F(3,45) = 3.49, P < 0.05$].

As was the case for the late positive response, the negativity elicited by ungrammatical relative to grammatical stimuli was larger for improbable than for probable items. Moreover, the grammaticality effect for probable items was largest over left anterior sites, while the grammaticality effect for improbable items was most evident at central sites, and slightly larger on the right [grammaticality \times probability \times hemisphere \times laterality: $F(1,15) = 4.71, P < 0.05$; grammaticality \times probability \times hemisphere \times laterality \times anteriority: $F(3,45) = 2.88, P < 0.05$]. The interaction between grammaticality and probability was reliable both for agreement [grammaticality \times probability: $F(1,15) = 5.34, P < 0.05$] and for pronoun case stimuli [grammaticality \times probability \times hemisphere \times laterality \times anteriority: $F(3,45) = 3.14, P < 0.05$].

These mean amplitude values for the 300–500 msec time window were also normalised according to the procedure outlined in McCarthy and Wood (1985) and subjected to repeated measures ANOVA with factors violation type (pronoun case *vs* verb agreement), grammaticality (grammatical *vs* ungrammatical), probability (improbable *vs* probable) and electrode site (26 sites). The scalp distribution of the negativity elicited in this time window still differed reliably as a function of violation type and of probability [violation type \times electrode site: $F(25,375) = 8.29, P < 0.001$; probability \times electrode site: $F(25,375) = 1.96, P < 0.05$].

Furthermore, violation type interacted with both grammaticality and probability, suggesting the distribution of grammaticality and probability effects differed in ERPs elicited by pronoun case stimuli and verb agreement stimuli [violation type \times grammaticality \times electrode site: $F(25,375) = 1.96, P < 0.05$; violation type \times probability \times electrode site: $F(25,375) = 2.06, P < 0.05$]. This observation was confirmed by *post-hoc* tests of individual violation types. Among the pronoun case stimuli, the distribution of the negativity was affected by grammaticality, but not probability [grammaticality \times electrode site: $F(25,375) = 2.15, P < 0.05$; probability \times electrode site: $F(25,375) = 0.91$]. However, among the verb agreement stimuli, the distribution of the negativity was affected by probability but not grammaticality [probability \times electrode site: $F(25,375) = 3.74, P < 0.01$; grammaticality \times electrode site: $F(25,375) = 0.97$].

DISCUSSION

Late Positivity

As expected from the results of previous investigations (Hagoort et al., 1993; Hagoort & Brown, 1994; Neville et al., 1991; Osterhout & Holcomb, 1992, 1993), we found the amplitude of the late positive response to linguistic

stimuli to be modulated by their grammaticality. However, we also found that the late positive response was modulated by the probability of encountering ungrammatical events within a given block of sentences (previously noted by investigators examining the ERP to Dutch materials: Gunter, Vos, & Mulder, 1995). Most importantly for present purposes, we obtained an interaction between the grammaticality and the probability effects on the amplitude of the late positive response. Overall, these results indicate that the late positivity elicited by syntactically anomalous stimuli cannot be viewed as a *direct* manifestation of a domain-specific parser, but rather a member of the P300 family of components (P3b and perhaps the slow wave) elicited by encountering the relatively rare linguistic event of ungrammaticality.

The identity of the P600/SPS and the P3b was further suggested by the similarity in the scalp distribution of the grammaticality and probability effects. Collapsed across grammaticality, the late positivity in the ERP to improbable stimuli displayed a remarkable similarity to the ERP to ungrammatical stimuli. In fact, the scalp distribution of the grammaticality effect was statistically indistinguishable from the distribution of the probability effect. Moreover, when we manipulate the local probability of encountering ungrammatical events so that the subject begins to expect them, we see a reduction in the amplitude of the late positivity elicited by ungrammatical words, and a slight enhancement in the positivity elicited by their grammatical counterparts.

Furthermore, our finding that the grammaticality effect was larger for the more salient pronoun case violations than the verb agreement errors, indicates the positivity may be more related to a reader's recognition of the ungrammatical stimulus as a rare event than to its grammaticality *per se*. Moreover, the modulation of the grammaticality effect by the probability manipulation—a manipulation which affects the participants' perception of ungrammaticality while leaving the facts of grammaticality unchanged—reinforces the extent to which the elicitation of the late positive response in the ERP is related to explicit perception of linguistic manipulations rather than to unconscious syntactic processing.

Our results differ from those of a similar experiment reported by Osterhout and colleagues (1996). Osterhout et al. jointly manipulated the frequency of subject–verb agreement errors and the presentation of verbs in upper- rather than lower-case letters to compare the probability effect for letter-case versus grammatical violations. Both types of violations appeared to elicit larger positivities when they were improbable; but, while the P300 elicited by infrequent letter-case violations was reliably larger than that elicited by frequent letter-case violations, a qualitatively similar modulation of the P600 elicited by grammatical violations did not reach significance. Interpreting the null result for grammatical violations, Osterhout et al. point

to the attention literature, arguing that probability manipulations of this magnitude (20 vs 60%) should have been sufficient to modulate the amplitude of the P300 component.

However, most experiments in the traditional P300 literature have employed stimuli such as pure tones and geometrical figures with little or no significance outside of the context of the laboratory. In contrast, natural language stimuli come from a familiar domain about which people have a rich set of expectations. One might surmise that the close correspondence traditionally observed between objective probability measures and the subjective probability registered by the experimental participant arises because people have few expectations about the artificial domains which have been used in the P300 literature. Because natural language stimuli come from a meaningful domain about which people have a rich set of priors, there is less reason to expect that the subjective probability indexed by the P3b will be yoked to the objective probability measure.

In addition, as grammatical violations go, subject-verb agreement errors are not particularly salient. In a violation-detection task for English sentences presented one word at a time, Kilborn (1988) reported an error rate of 33% for verb number agreement violations. Interestingly, in the subject-verb agreement condition in Kutas and Hillyard (1983), the largest positivity was seen for the participant who was a professional typesetter. Moreover, participants in Osterhout's experiments classified one out of 10 ungrammatical stimuli as acceptable—that is, they failed to notice the grammatical violation (Osterhout et al., 1996). These data suggest that experimental manipulation of the probability of encountering an ungrammatical sentence has to be quite dramatic to affect participants' perception of that probability. And, indeed, by varying the probability of grammaticality from 80 to 20%, and including the more salient pronoun case violations, we were able to detect a reliable effect of probability on the amplitude of the late positivity in participants' ERPs.

Note that we, like Osterhout et al. (1996), have expressed the probability manipulation in terms which are meaningful to us as experimenters: in terms of the number of sentences that contain either an ungrammatical verb or an ungrammatical pronoun. However, we have no reason to believe that our participants understood the experiment in exactly the same way we did. The participants may have tracked the relative proportion of ungrammatical versus grammatical sentences, ungrammatical versus grammatical clauses, ungrammatical versus grammatical verbs and pronouns, words which seem 'wrong' versus words which seem 'right', words which are hard to understand versus words which are easy, or any number of other possibilities. However, the ungrammatical events in this study were not random, but fell into two categories which participants might have been expected to detect.

While our participants might not have described the stimuli as subject-verb agreement and pronoun case violations, we do have reason to believe they were sensitive to the regularities and irregularities in the stimulus set. For example, during debriefing, the participants noted that they had read ungrammatical sentences and many noted that the proportion of ungrammatical sentences seemed to go up (or down) after the break between blocks. Furthermore, all participants were able to give examples of at least one of our two violation types (especially pronoun case violations), suggesting they had some awareness of the nature of the materials. Moreover, just as the grammaticality effect indicates the participants were sensitive to the morphosyntactic violations, the reliable probability effect suggests they did indeed notice our manipulation of the within-block probability of encountering a word which made the sentence ungrammatical.

We have argued that the sensitivity of the P600 to the probability of encountering an ungrammatical sentence in a block of trials suggests its membership of the P300 family of components. Alternatively, one might argue that the observed probability effects are nonetheless consistent with the operation of a syntax-specific processing mechanism. For example, if the P600 indexes the diagnosis of a bad parse and the need for syntactic reanalysis, then repeated presentation of violation types might facilitate this diagnosis process and, consequently, lead to a decrease in the amplitude of the P600 component. Similarly, if the P600 indexes the reanalysis process itself, one might well expect decreased P600 amplitude due to the fact that repetition of violation types makes syntactic reanalysis easier.

However, this sort of account does not explain our observation of an increase in the amplitude of the positivity elicited by grammatical stimuli in grammaticality improbable blocks. Even if the probability manipulation makes the recognition of ungrammatical stimuli easier, there is no *prima facie* reason it should make the recognition of grammatical stimuli more difficult. Similarly, if the probability manipulation facilitates syntactic reanalysis of the ungrammatical stimuli, why would it be elicited by grammatical stimuli that do not require reanalysis?

We suggest that the increased amplitude of the late positivity elicited by grammatical stimuli in grammaticality improbable blocks relative to identical stimuli in grammaticality probable blocks reflects the way that participants update the contextual models which govern their expectations. Although participants' default expectation is for sentences to be grammatical, the high probability of encountering ungrammatical sentences in grammaticality improbable blocks might lead them to expect sentences to be ungrammatical in particular ways.

Because the P3b is an index of subjective rather than objective probability, its elicitation depends on the extent to which the objective

manipulation is salient to experimental participants. Besides subjective probability, the amplitude of the P3b is also modulated by the task-relevance of the eliciting event (Duncan-Johnson & Donchin, 1977). These and other findings have led Donchin (1981; Donchin & Coles, 1988) to postulate the context-updating model of the P300. On Donchin's model, the P300 indexes processes involved in updating an individual's mental model of the environment. Thus improbable events elicit P3b not because they are unexpected *per se*, but because they prompt the participant to update her model of the environment.⁴ Accordingly, grammatical events in the ungrammaticality probable blocks elicit P3b because the probability manipulation leads participants to change their contextual models.

However, researchers sceptical of our hypothesis that the late positivity in ERPs elicited by morphosyntactic violations is an instance of an ERP response elicited by unexpected events in general, might still make recourse to an alternative explanation. One might argue that the probability manipulation was not equally salient in the case of the grammatical and the ungrammatical events. After all, ungrammatical events rendered the whole sentence ungrammatical, while grammatical events merely maintained the grammatical *status quo*. In this case, we might expect the probability manipulation to have a larger effect on the ungrammatical events than on the grammatical controls.

Of course, by admitting that the context-updating mechanism might be differentially sensitive to grammatical and ungrammatical events, the sceptic has moved a long way towards our position. Nonetheless, let us explore this alternative explanation of the observed pattern of data. Assume p_1 equals the probability effect for grammatical events, p_2 equals the probability effect for ungrammatical events, and that p_2 is greater than p_1 . For the sake of argument, we assume that our experimenter defined probability figures of 80 and 20% are correct. As P3 amplitude is (generally) inversely proportional to the probability, the amplitude of probable events will be a multiple of 1.25, while the amplitude of improbable events will be a multiple of 5. Finally, the amplitude of the syntax-specific ERP component is equal to g .

If the amplitude of the grammaticality effect were independent of the probability effect, the amplitude of the late positivity in each of our four subconditions might be described as follows:

$$\text{Ungrammatical probable: } P3/P6 = g + 1.25*p_2$$

$$\text{Grammatical probable: } P3/P6 = 0*g + 1.25*p_1$$

$$\text{Ungrammatical improbable: } P3/P6 = g + 5*p_2$$

⁴ Thus the elicitation of a P3b does not require one to actively expect a particular stimulus to occur at a given time.

Grammatical improbable: $P3/P6 = 0 * g + 5 * p_2$

The grammaticality effect on the probable stimuli would be described by the linear function $P6 = 1.25 * (p_2 - p_1) + g$, while the grammaticality effect on the improbable stimuli would be described by a similar function with a steeper slope: $P3 = 5 * (p_2 - p_1) + g$. The alternative hypothesis is in principle consistent with the observed pattern of data, as well as with the independent generation of the P600/SPS and the P3b.

However, this sceptical hypothesis implies that the amplitude of the putatively syntax-specific late positivity also reflects activity due to context updating, or some other probabilistic process that weighs grammatical and ungrammatical stimuli differently. Moreover, one need not explicitly manipulate probability to observe the effects of the probabilistic process on the late positivity. The same problem obtains for the more typical case in which ungrammatical sentences comprise half of the sentence set. In such a case, the amplitude factor of the probability parameter would be 0.5 for both the grammatical and the ungrammatical stimuli. However, our sceptic's previously invoked parameters p_1 and p_2 remain unequal, so that the amplitude of the grammaticality effect is described by the function $P6 = 0.5(p_1 - p_2) + g$.

Thus by invoking free parameters, the sceptic can explain away the observed pattern of effects. But in so doing, she is forced to forego hope of ever finding an accurate index of the true grammaticality effect g . While in another circumstance we might try to disentangle grammaticality and probability by their scalp distributions, the two effects cannot be so distinguished (a fact which the sceptic presumably attributes to an unlucky coincidence). It would seem that the sceptic has done an uncanny job of vaporising the baby along with the bathwater.

In contrast, we argue that the original discovery of a grammaticality effect on the amplitude of the late positivity resulted from the fact that people perceive actual and apparent grammatical violations as a deviation from the normal course of events. This in turn prompts them to update their contextual models of the linguistic environment. The similarity in the scalp distributions of the positivity elicited by ungrammatical events and the positivity elicited by improbable events is not coincidental, but arises because the same neural generators contribute to both effects. Consequently, variables such as odd-ball probability and event-saliency, known to influence the amplitude of the P3b component, can be seen to affect the amplitude elicited by morphosyntactic violations in predictable ways.

Researchers who would prefer to interpret these results strictly in terms of the violation of grammatical rules must explain why the probability that a rule had been violated had a non-additive effect on the amplitude of the

P600/SPS. Likewise, they will have to account for the effect of the saliency of the error type on P600/SPS amplitude. The P3b, on the other hand, is well-known to be affected in this way by variables such as probability and saliency. Thus, the hypothesis that the P3b and the P600/SPS are essentially identical is more parsimonious.

Perhaps because a lifetime of language experience makes ungrammatical stimuli *prima facie* improbable, experimental manipulation of the probability of encountering ungrammatical sentences is more salient than the equivalent manipulation of probability of grammatical sentences. Note, however, that even grammatical stimuli elicit enhanced late positivities when they occur in a block in which 80% of the sentences are ungrammatical. These data suggest the positive shift elicited by morphosyntactic violation is an instance of a domain-general cognitive process. Nonetheless, anterior negative effects may index processes which differentiate linguistic violations from the detection of improbable events in general.

Anterior Negative Effects

Besides late positivities, ERPs to syntactic violations have often included left lateralised anterior negativities (LAN) between 300 and 500 msec post-stimulus (Neville et al., 1991; Osterhout & Holcomb, 1992; Rösler et al., 1993). Although these negative effects in the ERP have received less attention than the P600/SPS, the LAN may be a more direct index of syntactic processing than the P600/SPS. Rösler et al. (1993) observed a LAN to violations of subcategorisation constraints in German and contrasted it with the N400 response elicited by semantic violations in another group of subjects.

Münste, Heinz and Mangun (1993) provided further evidence of the distinction between the LAN and the N400 by using a more extensive array of electrodes as well as a within-subjects comparison of ERPs to the second word of a word pair which either was or was not semantically related or grammatically correct. Semantically incorrect words elicited the classic N400 effect with a centro-parietal maximum. In contrast, the syntactic anomalies elicited a LAN whose scalp distribution was reliably different from that of the N400, and whose onset was 50 msec later. Münste et al. (1993) argue that the LAN effects observed in their study are distinct from the N400 component, produced by different underlying generators, and index syntactic aspects of processing.

Kluender and Kutas (1993a, 1993b), however, have suggested that the LAN elicited by certain syntactic violations actually indexes some aspect of working memory usage. Kluender and Kutas found a consistent LAN effect associated with entering a filler in working memory, storing it and

subsequently retrieving it to assign fillers to gaps. This is consistent with data reported by King and Kutas (1995), who found a similar LAN effect in ERPs elicited by the verb immediately following the gap in object-relative clauses which tax working memory in comparison to the corresponding verb in subject-relative clauses (e.g. “admitted” in the examples below), and a less lateralised anterior negativity difference between subject-relative clause and non-relative clause control sentences.

Subject relative

9. The reporter who harshly attacked the senator admitted the error.

Object relative

10. The reporter who the senator harshly attacked admitted the error.

Furthermore, Shao and Neville (1996) observed LAN effects elicited by two sorts of *semantic* violations, namely hyponymy (e.g. Jane never eats any *meat* so instead she eats lots of *beef*) and negative polarity (e.g. Cliff believes that he has *ever* seen that woman before). Shao and Neville’s findings are at least amenable to a *post-hoc* working memory account. For instance, the hyponymy violation might induce readers to refer back to the category “meat” to confirm the occurrence of the violation. The negative polarity violation might induce readers to (vainly) search working memory for the occurrence of the negative element which would license the negative polarity item “ever”.

Our results suggest the negativities elicited by pronoun case stimuli differed qualitatively from those elicited by verb agreement stimuli. Whereas the distribution of the ERP elicited by the pronouns was modulated by the *grammaticality* of the stimulus, the distribution of the ERP elicited by the verbs was modulated by the *probability* manipulation. Moreover, the distribution of the negativity elicited by the pronoun case stimuli was anterior, much like the LAN reported by previous investigators (Müntz et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1992; Rösler et al., 1993); while the distribution of the negativity elicited by verb agreement stimuli was more centro-parietal, consistent with the distribution of both the N2 and the N400 components. Although the elicitation of the LAN in this study was unexpected, its congruity with previous reports render it unsurprising.

Besides having the effect of making the sentences ungrammatical, our “grammaticality” manipulations also affected the referentiality and interpretability of the stimuli in ways which may have altered their processing difficulty. Although English is morphologically impoverished compared with languages that allow for more word order variation, the pronoun case system is one of its more salient aspects. Our intuition is that

verb agreement errors in English produce a subtle change in the sentence meaning, given the relatively low availability of agreement information in the language and the reports that agreement information is virtually ignored in certain experimental tasks with English-speaking subjects (e.g. MacWhinney, Bates, & Kliegl, 1984). Pronoun case violations, however, seem to have a more drastic effect due to the relationship between pronoun case marking and the assignment of thematic roles such as agent or patient. In particular, the processing of pronouns recruits working memory resources to recover their referents, to track referential relationships and to add element(s) to the discourse model based on contextual information and background knowledge. Consequently, the pronoun case violations we employed may have invoked working memory operations to a greater extent than did the verb agreement violations, as the participants attempted to establish sensible co-reference with existing or novel entities in the discourse model.

This argument provides an independent reason why grammaticality (but not probability) could modulate the amplitude of the anterior negativity (LAN) elicited by the pronoun case violations. These data suggest the LAN effect, unlike the P600/SPS, may index processes which differentiate language processing from the detection of anomalous events.⁵ In particular, we have noted the compatibility of our results with Kluender and Kutas' (1993a, 1993b) suggestion that the LAN indexes some aspect of working memory usage. Although the results in the literature do not necessarily imply the LAN indexes cognitive operations which are exclusively *syntactic*, this ERP component may index operations specific to verbal working memory.

ERPs and Domain Specificity

One argument made by researchers taken with the syntactic character of the anomalies which elicit the P600/SPS concerns its sensitivity to syntactic but not semantic manipulations (Hagoort et al., 1993; Hagoort & Brown, 1994; Osterhout & Holcomb, 1992, 1993). Regardless of the relationship between

⁵ Münte, Matzke and Johannes (1997) have shown that LAN effects elicited by verb agreement violations in German are accompanied by SPS/P3b effects when the violations occur in normal prose, but not when the violations occur in sentences where content words have been replaced by (appropriately case-marked) pseudowords. In this case, we would argue that no SPS is elicited because the saliency contrast of the syntactic violations in pseudoprose compared to "correct" pseudoprose is very low. The LAN effect, however, is a possible index of morphosyntactic processing (but see text for further elaboration). We note in passing that the absence of SPS/P3b in German pseudoprose, the small SPS/P3b in Dutch syntactic prose (Hagoort & Brown, 1994) and the robust SPS/P3b reported herein to English prose, suggest a possible relationship between SPS/P3b amplitude and the continuum of linguistic (and paralinguistic) contexts in which one might notice irregular morphology.

the neural generators of the P600/SPS and the anatomical locus of the parser, it behooves proponents of interactive approaches to explain why the P600/SPS is elicited by syntactic violations but not by contextually inappropriate words. In a similar vein, why do syntactic violations elicit an increased LAN response while contextually inappropriate words primarily modulate the N400 component?

For example, Hagoort et al. (1993) contend that the dissociation between the sensitivity of the N400 and the P600/SPS to semantic and syntactic information, respectively, argues for distinct neurophysiological “signatures” for processing the two sorts of information:

... the processing of syntactic information has a neurophysiological signature that is clearly different from that for the processing of semantic information. This result is difficult to account for in sentence processing models that deny that qualitatively different constraints (i.e., syntactic and semantic) make qualitatively different contributions to the construction of an interpretation of the whole utterance (e.g., McClelland, St.John, & Taraban).

However, it is unclear whether sentence processing models alluded to by Hagoort et al. (1993) even exist. While statements may have been made by certain investigators a-flush in the afterglow of a paradigm shift, we are aware of no serious sentence processing models which equate syntactic and semantic constraints. In any case, the questions of interest concern first, whether there is a level of representation which is completely independent of meaning, and, second, whether and when different sorts of linguistic constraints interact in the course of language processing.

Our own position is *not* that syntactic and semantic constraints are indistinguishable, but rather that the division of symbolic units into separable components is arbitrary. This position is motivated by functional linguistic theories, such as Cognitive Grammar (Langacker, 1987) and Construction Grammar (Fillmore, Kay, & O'Connor, 1988), which suggest syntactic descriptions can be reduced to pairings between forms and meanings. Rather than assuming grammar is an autonomous level of linguistic description, cognitive grammar claims that the lexicon, morphology and syntax form a continuum of symbolic structures in which traditional syntactic structures are the most schematic.

One interesting difference between syntactic and semantic violations is the extent to which the former appear to be more *categorical* in nature. That is, native speakers can easily classify sentences as either acceptable or unacceptable based on naive theories of language usage. By contrast, semantic violations historically employed in the ERP literature (e.g. Kutas & Hillyard, 1980) are less amenable to classification as rare events. Although there are statistical regularities in the co-occurrence of semantically related

lexical items, pragmatic phenomena such as metaphor and irony reveal speakers to be quite adept at interpreting words in seemingly inappropriate contexts. Consequently, readers may be less apt to spontaneously classify sentences such as “The pizza was too hot to cry” as categorically unacceptable.

However, this line of reasoning predicts that semantic anomalies should also elicit late positivities when those anomalies are easily classifiable and/or task-relevant. To wit, Shao and Neville (1996) report centro-parietal positivities in ERPs elicited by three types of semantic anomalies, including violations of hyponymy, negative polarity and contextually inappropriate words similar to those employed by Kutas and Hillyard (1980). Kutas (1993) reports a late positivity in the second word of semantically related word pairs. Also, Schwartz et al. (1996) found late frontal positivities in a category priming task in the elderly. The significance of this latter result stems from the unconfounding effect ageing has on the overlapping N400/P3b component problem. Elderly participants generate reduced amplitude N400s (e.g. Iragui, Kutas, Mitchiner, & Hillyard, 1993), thus minimising the obscuring effect of this preceding negativity, and revealing the late positivity. Overall, claims that all and/or only syntactic anomalies elicit late positivities are (at best) overstated and at worst spurious.

Another argument against the identity of the P600/SPS and the P3b has involved the issue of task relevance. For example, Hagoort et al. (1993) maintain that their elicitation of the P600/SPS in the absence of an explicit task argues against the identity of the two components. Similarly, Osterhout et al. (1996) argue that task relevance does not modulate P600 on the basis of a non-significant difference in the P600 amplitude in a group instructed to read sentences and a group explicitly instructed to make grammaticality judgements.

However, these arguments do not adequately consider the extent to which participants in a psycholinguistics experiment will naturally consider the grammaticality of stimuli as a relevant factor. Moreover, attention-grabbing stimuli elicit P300 regardless of task relevance; even a no-task situation elicits P300 in contrast to a condition where stimuli are being actively ignored. A strong test of whether P600/SPS amplitude is modulated by task relevance would involve a comparison between a task which directs participants' attention away from grammaticality versus a task which accentuates grammaticality as a task-relevant stimulus dimension.

The identification of the P600/SPS with the P3b suggests that the term SPS is a bit of a misnomer. The P3b elicited by syntactic anomalies is domain-general and not specifically tied to syntactic processing, or even to language in general. Moreover, identification of the P600/SPS with the P300 family of components clears up the mystery as to why the purported index of syntactic processing occurs *after* semantic information is available. Because

it is commonly assumed that semantic and pragmatic processing depend upon the prior output of the parser, the latency of the P600/SPS has never sat well with the presumed timecourse of sentence processing. Because results from the semantic priming literature suggest semantic information is available within 200 msec of word onset, one would expect a direct index of syntactic processing to be manifest much earlier than the 500 msec after word onset of the P600.

However, the identification of the P600/SPS as an ERP component which indexes the conscious recognition of morphosyntactic violations should come as no surprise. After all, the conscious grammaticality judgement is the linguist's most important empirical tool. All native speakers can make these judgements, including children (Bates & MacWhinney, 1989) and even agrammatic aphasics (Linebarger, Schwartz, & Saffran, 1983). Osterhout and Holcomb (1992) point out that regardless of whether the P600 is a member of the P300 family, an ERP component sensitive to the syntactic well-formedness is quite useful to psycholinguists: "if the P600 is shown to occur reliably whenever comprehenders are presumed to encounter syntactic anomaly, then the P600 is a useful tool even in the absence of certain knowledge concerning the cognitive events underlying the component".

Because it is reliably elicited by syntactic anomalies, the P600/P3b is a useful dependent variable in psycholinguistic research. Because its amplitude varies with the salience of syntactic anomaly, and its latency varies with the complexity of processing, the P600/P3b can be fruitfully employed to test hypotheses about the salience and/or complexity of various types of syntactic anomalies. Furthermore, given that ERP effects in the P300 family tend to be large compared to those in other endogenous components, reliable ERP effects might be observed for a relatively low number of stimuli per condition. Statistical power associated with the P3b enables testing hypotheses about the nature of syntactic ambiguity resolution (see MacDonald et al., 1994, for a review of this issue). The P600/P3b could also serve as a good dependent variable for testing hypotheses such as those in the competition model (Bates & MacWhinney, 1989) about the relative importance of linguistic cues such as animacy, word order and morphology.

Although physiological measures do not appear to bear an unambiguous relationship to linguistic variables apart from the rest of the cognitive system, ERP measures are quite sensitive relative to reaction times in their ability to discriminate different types of processing. However, a few caveats are in order, as it is important to consider variables which influence the P300 components when using the P600/P3b as a dependent measure in psycholinguistic experiments. The P600/P3b elicited by particular syntactic anomalies is dependent upon the composition of the entire stimulus set.

That is to say, the P600 elicited by a particular subcategorisation violation will vary as a function of the stimulus set in which it is embedded. For example, if it occurs in the context of normal prose, the subcategorisation violation will probably seem more salient than if it occurs in the context of other syntactic violations.

Results demonstrate that the amplitude of the late positivity formerly associated only with syntactic anomaly also varies as a function of the proportion of grammatical to ungrammatical sentences. In particular, the non-additivity of grammaticality and probability on the late positive ERP response argues against the hypothesis that the P600/SPS to morphosyntactic errors and the P3b associated with domain-general processing are generated independently. Overall, the results argue against an account of the P600/SPS as the index of a domain-specific parser, and for an account of the P600 as the index of a domain-general mechanism.

Manuscript received July 1996

Revised manuscript received April 1997

REFERENCES

- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds), *The crosslinguistic study of sentence processing*, pp. 3–76. Cambridge: Cambridge University Press.
- Bloom, P.A., & Fischler, I.S. (1980). Completion norms for 329 sentence contexts. *Memory and Cognition*, 8, 631–642.
- Coles, M.G.H., & Rugg, M.D. (Eds) (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press.
- Coles, M.G., Gratton, G., & Fabiani, M. (1990). Event-related brain potentials. In J.T. Cacioppo & L.G. Tassinary (Eds), *Principles of psychophysiology: Physical, social, and inferential elements*, pp. 413–455. Cambridge: Cambridge University Press.
- Dale, A. (1994). *Source localization and spatial discriminant analysis: Linear approaches*. PhD dissertation, University of California San Diego, La Jolla, CA.
- Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked brain potentials and behavior*, pp. 13–88. New York: Plenum Press.
- 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, 357–374.
- Duncan-Johnson, C.C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 14, 456–467.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.
- Fillmore, C.J., Kay, P., & O'Connor, M.C. (1988). Regularity and automaticity in grammatical constructions: The case of let alone. *Language*, 501–538.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. PhD dissertation, University of Connecticut, Storrs, CT, distributed by the Indiana University Linguistics Club.

- Gunter, T.C., Vos, S.H., & Mulder, G. (1995). Syntactic violations and ERPs: P600 or P3b? Poster presentation at the *CUNY Sentence Processing Conference*, Tucson, AZ, March.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton Jr, L. Frazier, & K. Rayner (Eds), *Perspectives on sentence processing*, pp. 45–80. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.
- Hillyard, S.A., & Picton, T.W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), *Handbook of physiology. Higher functions of the nervous system. Section 1: The nervous system. Vol. V: Higher functions of the brain, Part 2*, pp. 519–584. Washington, DC: American Physiological Society.
- Huynh, H., & Feldt, L.S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82.
- Iragui, V.J., Kutas, M., Mitchiner, M.R., & Hillyard, S.A. (1993). Effects of aging on event-related potentials and reaction times in an auditory discrimination task. *Psychophysiology*, 30, 10–22.
- Johnson, R., Jr (1986). A triarchic model of P300 amplitude. *Psychophysiology*, 23, 367–384.
- Johnson, R., Jr (1989). Auditory and visual P300s in temporal lobectomy patients: Evidence for modality-dependent generators. *Psychophysiology*, 26, 633–650.
- Kilborn, K. (1988). *A note on the on-line nature of discourse processing: Comparing behavioral and electrophysiological measures in monitoring for semantic anomalies and grammatical errors*. Max-Planck Institute for Psycholinguistics Technical Report. Nijmegen: Max-Planck-Institute for Psycholinguistics.
- King, J.W., & Kutas, M. (1995). Who did what and when? Using word- and clause-related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 378–397.
- Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196–214.
- Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573–633.
- 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. (1980). 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., & Kluender, R. (1994). What is who violating: A reconsideration of linguistic violations in light of event-related brain potentials. In H.J. Heinze, T.F. Münte, & G.R. Mangun (Eds), *Cognitive electrophysiology*, pp. 183–210. La Jolla, CA: Birkhauser Boston.
- Kutas, M., Lindamood, T.E., & Hillyard, S.A. (1984). Word expectancy and event-related brain potentials during sentence processing. In E.S. Kornblum & J. Requin (Eds), *Preparatory states and processes*, pp. 217–237. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, 69, 218–233.
- Langacker, R.W. (1987). *Foundations of cognitive grammar, Vol. 1: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Linebarger, M.C., Schwartz, M.F., & Saffran, E.M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13, 361–392.

- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
- MacWhinney, B., Bates, E., Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.
- McCarthy, G., & Wood, C.C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- Münte, T.F., Heinze, H.J., & Mangun, G.R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, *5*, 335–344.
- Münte, T.F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 300–311.
- Neville, H., Nicol, J.L., 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.
- Osterhout, L. (1990). *Event-related brain potentials elicited during sentence comprehension*. Unpublished doctoral dissertation, Tufts University, Medford, MA.
- 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*, pp. 15–44. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Osterhout, L., & Holcomb, P. (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*, 413–437.
- Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, *8*, 507–526.
- Picton, T.W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–479.
- Pritchard, W.S. (1981). Psychophysiology of P300: A review. *Psychological Bulletin*, *89*, 506–540.
- Rösler, F., Pütz, P., Friederici, A.D., & Hahne, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of Cognitive Neuroscience*, *5*, 345–362.
- Rugg, M.D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition*, *18*, 367–379.
- Schwartz, T.J., Kutas, M., Salmon, D.P., Paulsen, J.S., & Butters, N. (1996). Electrophysiological insights into the nature of the semantic deficit in Alzheimer's disease. *Neuropsychologia*, *34*, 827–841.
- Shao, J., & Neville, H.J. (1996). ERPs elicited by semantic anomalies: Beyond the N400. In *Proceedings of the Third Meeting of the Cognitive Neuroscience Society*, pp. 73–73, San Francisco, CA, March.
- Spivey-Knowlton, M., & Tanenhaus, M.K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton Jr, L. Frazier, & K. Rayner (Eds), *Perspectives on sentence processing*, pp. 415–439. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*, 380–393.
- Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, *11*, 343–356.