PERGAMON

Journal of Neurolinguistics 15 (2002) 171–187

Journal of NEUROLINGUISTICS

www.elsevier.com/locate/jneuroling

Words in the brain: lexical determinants of word-induced brain activity

Lee Osterhout*, Mark Allen, Judith McLaughlin

Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195, USA

Abstract

Many studies have shown that open- and closed-class words elicit different patterns of brain activity, as manifested in the scalp-recorded event-related potential (ERP). One hypothesis is that these ERP differences reflect the different linguistic functions of the two vocabularies. We tested this hypothesis against the possibility that the word-class effects are attributable to quantitative differences in word length. We recorded ERPs from 13 scalp sites while participants read a short essay. Some participants made sentence-acceptability judgments at the end of each sentence, whereas others read for comprehension without an additional task. ERPs were averaged as a function of word class (open versus closed), grammatical category (articles, nouns, verbs, etc.), and word length. Although the two word classes did elicit distinct ERPs, all of these differences were highly correlated with word length. We conclude that ERP differences between open- and closed-class words are primarily due to quantitative differences in word length rather than to qualitative differences in linguistic function.

Keywords: Event-related potentials; Grammatical categories; Language comprehension

1. Introduction

Is the brain response to words determined primarily by their linguistic functions or by their physical forms? This question is at the center of a debate concerning the biological status of two lexical categories, the open- and closed-class vocabularies. The distinction between these two vocabularies roughly parallels the distinction between syntax (sentence form) and semantics (sentence meaning). Closed-class words (e.g. articles, auxiliary verbs, prepositions, and pronouns) typically act as agents of phrasal construction, whereas open-class words (e.g. nouns, verbs, and adjectives) convey meaning by referring to specific objects and events. Recent evidence has been taken to indicate that these two vocabularies elicit qualitatively distinct brain responses, as reflected in scalp-recorded

* Corresponding author. Tel.: +1-206-543-4177; fax: +1-206-685-3157.
E-mail address: losterho@u.washington.edu (L. Osterhout).

0911-6044/02/$ - see front matter
PII: S0911-6044(01)00036-7

event-related brain potentials (ERPs). However, it is currently unclear whether these differences in brain response to open- and closed-class words are due to differences in the linguistic function of these words, or to differences in physical form. We show here that although the two word classes do elicit quite different ERPs, these differences are in fact primarily a function of the physical, rather than linguistic, properties of words.

A variety of evidence seemingly reflects the functional and biological differences between open- and closed-class words. For example, studies of aphasic syndromes have suggested that anterior lesions in the left hemisphere selectively disrupt the use of the production and comprehension of closed-class words, whereas lesions to posterior regions disrupt use of open-class words (Caramazza & Zurif, 1976; Damasio & Damasio, 1989; Heilman & Scholes, 1976; Saffran, Schwartz, & Marin, 1980; Schwartz, Saffran, & Marin, 1980). Research with normal adults has indicated that the two word classes participate in distinct types of speech errors (Garrett, 1980) and display different patterns of hemispheric asymmetries (Bradley & Garrett, 1983). Such evidence has been taken to indicate that open- and closed-class words might be both functionally and neuroanatomically distinct.

A particularly compelling type of evidence would be neurobiological evidence from the normal brain that the two word classes are distinct. One means for obtaining such evidence involves recording ERPs elicited during language processing. ERPs are a continuous, on-line record of the brain's electrical activity that provides at least rough estimates of lateralization and localization of function (cf. Coles & Rugg, 1995). ERPs consist of positive and negative voltage peaks (or 'components') that are distributed over time. It is generally agreed that ERP components with distinct scalp distributions are necessarily generated by distinct neural systems (cf. Johnson, 1993). Therefore, evidence that open- and closed-class words elicit different ERP components, or different distributions of components across the scalp, would support the notion that the two word classes are neurally (and, by extension, cognitively) distinct.

Evidence of exactly this type has been reported by Neville, Mills, and Lawson (1992). These researchers recorded ERPs while participants read a set of unrelated sentences. Neville et al. reported several differences in the ERPs to open- and closed-class words. Importantly, the two word classes elicited negative components that clearly differed in their temporal and spatial properties. Closed-class words elicited a negative component that was largest over anterior regions of the left hemisphere and that peaked at about 280 ms (*N280*). By contrast, open-class words elicited a large amplitude, posteriorly distributed negativity that peaked between 350 and 400 ms (*N400*). The distribution of these negativities corresponds well with the clinical observation that left anterior lesions disrupt use of closed-class words whereas posterior lesions disrupt use of open-class words.[1]

These results seem to provide compelling support for the hypothesis that qualitative differences exist in the representation and function of open- and closed-class words, and that the functional differences between these categories underlie dramatic differences in

[1] Closed-class words also elicited an N400 component, albeit much smaller in amplitude than the N400 component elicited by open-class words. The fact that both word classes elicit an N400 might be viewed as problematic for the notion that the N280/N400 distinction reflects the accessing of different lexicons for the two word classes.

brain activity. This conclusion is tempered by two caveats. First, the numerous studies contrasting the ERP responses to open- and closed-class words have produced an amazingly varied set of results (for a summary, see Brown, Hagoort & ter Keurs, 1999). Not all of these results are consistent with those reported by Neville et al. For example, not all researchers have reported an N280-like response to closed-class words (Kutas & Hillyard, 1983; Kutas, Van Petten, & Besson, 1988; Münte et al., 2001; Osterhout, Bersick, & McKinnon, 1997; Pulvermüller, Lutzenberger, & Birbaumer, 1995), and others have found that both word classes elicit similar, rather than dissimilar, negative-going components (Brown et al., 1999; Garnsey, 1985; Nobre & McCarthy, 1994; ter Keurs, Brown, Hagoort, & Stegeman, 1999; Van Petten & Kutas, 1991). These disparate results might be due in part to wide variation across experiments in the type of stimuli presented and in the task assigned to participants. Stimuli have included prose (Kutas & Hillyard, 1983), a series of unrelated sentences (Neville et al., 1992; Van Petten & Kutas, 1991), and lists of unrelated words (Garnsey, 1985; Nobre & McCarthy, 1994; Pulvermüller et al., 1995). Tasks have included reading for comprehension (Kutas & Hillyard, 1983), sentence-acceptability judgments (Neville et al., 1992; Osterhout et al., 1997), and lexical decisions (Pulvermüller et al., 1995). Despite these varying conditions, other word-class effects have been consistently observed. For example, many researchers have reported that closed-class words elicit a ramp-like negativity between about 400 and 700 ms after presentation of the word (labeled the *N400–700* effect). Several researchers have proposed that this ERP difference, rather than the N280/N400 distinction, reflects the different linguistic functions of the two word classes (Brown et al., 1999; Münte et al., 2001; ter Keurs et al., 1999; van Petten & Kutas, 1991).

A second reason for caution in interpreting the Neville et al. results (and any other ERP difference between word classes) revolves around the difficulty of determining which aspects of the vocabulary distinction (word length, frequency of usage, repetition, contextual constraints, abstractness of meaning, referentiality, grammatical role, etc.) are responsible for the reported ERP differences (cf. Kutas & Van Petten, 1994; Osterhout et al., 1997). In particular, it is critical that quantitative differences in the physical properties of words (e.g. word frequency and length) are not mistaken for qualitative differences in linguistic function. Although some studies have carefully controlled for or measured the factors of word frequency and length (Brown, et al., 1999; Garnsey, 1985; Osterhout et al., 1997; Pulvermüller et al., 1995; ter Keurs et al., 1999), others have not (e.g. Kutas & Hillyard, 1983; Van Petten & Kutas, 1991). Neville et al. examined the effects of word frequency and length by forming separate averages for high and low frequency and short and long open-class words. None of these categories elicited an N280-like effect, but all categories elicited an N400. Furthermore, the primary effect of word frequency was to alter N400 amplitude (such that less-frequent words elicited larger N400s; see also Van Petten & Kutas, 1991), whereas the major effect of word length was to add a symmetrical frontal positivity in the ERPs to short words, beginning at about 200 ms and continuing for several hundred ms. Based on these results, Neville et al. concluded that the N280/N400 distinction was in fact a function of word class rather than word frequency or length. Other researchers have also failed to find evidence of frequency or length effects (Brown et al., 1999; ter Keurs et al., 1999) that might account for ERP differences between word classes.

These null results can be contrasted with two recent studies reporting compelling length and frequency effects. Osterhout et al. (1997) formed averages over grammatical categories such as determiners, pronouns, nouns and verbs. All categories elicited a negativity that peaked between 280 and 400 ms. The latency of the negativities was regressed onto the mean normative frequency and mean length of words in each category. Length and frequency were highly correlated with the latency of the negativity, and accounted for more than 80% of the latency variance in most subjects. Similarly, King and Kutas (1998) reported that although both closed- and open-class words elicited a 'lexical processing negativity' over anterior sites in the left hemisphere, the latency of this effect was a linear function of both word frequency and length.

Clearly, no consensus exists as to whether or not open- and closed-class words elicit distinct ERP effects; whether these effects (if they exist) reflect quantitative differences in physical properties of words or qualitative differences in linguistic function; and how variations in the tasks assigned to subjects might influence the results. In the study reported here, we carefully examined the brain responses to words presented in a short prose passage. ERPs were averaged as a function of word class, grammatical category, and length in letters. Subjects were asked to either read the prose for comprehension and also to make sentence acceptability judgments at the end of each sentence or to simply read for comprehension. We report that most of the robust differences in the ERPs elicited by open- and closed-class word can be accounted for by variations in word length.

2. Method

2.1. Subjects

Twenty right-handed native-English speakers participated for class credit or for a small monetary compensation. Ages ranged from 18 to 45 (mean = 28) years. Ten subjects participated in each of the task conditions.

2.2. Materials

A short essay describing Amelia Earhart's last flight was selected for presentation. This text was a modified version of an essay taken from an English as a Second Language text (Kenan, 1986). The text contained 173 sentences that ranged from four to 15 words in length. For 60 of these sentences, a sentence-embedded word was replaced with either a semantically anomalous (30 sentences) or syntactically anomalous (30 sentences) word from the same grammatical category. The ERP responses to these linguistic anomalies are reported elsewhere (Osterhout, McLaughlin, and Allen, in preparation). ERPs to the syntactically and semantically anomalous words, and from words that immediately followed the anomalous words, were not included in the analyses reported below. Sentence-initial and sentence-final words were also excluded in the grand average wave-forms. This was done to avoid confounding the effects of word class with presumed processes that are specific to the sentence-initial and sentence-final positions (e.g. sentence wrap-up processes). The remaining text contained 742 closed-class words (including 121 articles, 130 prepositions, 93 pronouns, and 146 auxiliary verbs) and 416 open-class words (including 170 nouns and 144 verbs).

2.3. Procedure

Participants sat comfortably in an isolated, dimly illuminated room. The text was presented as a series of sentences. Each sentence trial consisted of the following events. A fixation cross appeared for 500 ms, after which a sentence was presented in a word-by-word manner, with each word appearing on the center of the screen for 300 ms. A blank-screen interval of 400 ms separated words. Sentence-ending words appeared with a period or other appropriate punctuation. A 1450-ms blank-screen interval followed each sentence and provided participants with an opportunity to blink and rest their eyes. This interval was followed by a prompt asking participants to respond by pressing a button on a joystick. The hand(s) used to respond was (were) counterbalanced across participants. The subjects' task was manipulated in a between-subjects manner. In one condition, subjects passively read the text for comprehension and, when presented with the response prompt, pushed one of the buttons when ready for the next sentence. In the second condition, subjects were asked to read for comprehension and to make sentence acceptability judgments at the end of each sentence. They indicated their response by pressing one button on the joystick for 'acceptable' and the other button for 'unacceptable.' Each session lasted approximately 2 h.

2.4. Data acquisition and analysis

Continuous EEG was recorded from 13 scalp sites using tin electrodes attached to an elastic cap (Electrocap International). Electrode placement included International 10–20 system locations (Jasper, 1958) over homologous positions over the left and right occipital (O1, O2) and frontal (F7, F8) regions and from frontal (Fz), central (Cz), and parietal (Pz) midline locations. In addition, several non-standard sites over posited language centers were used, including Wernicke's area and its right hemisphere homologue (WL, WR: 30% of the interaural distance lateral to a point 13% of the nasion–inion distance posterior to Cz), posterior temporal (TL, TR: 33% of the interaural distance lateral to Cz), and anterior temporal locations (ATL, ATR: one-half the distance between F7/F8 and T3/T4). Vertical and horizontal eye movements were monitored by means of two electrodes, placed beneath the left eye and to the right of the right eye, respectively. The above 15 channels were referenced to an electrode placed over the left mastoid bone and were amplified with a bandpass of 0.01–100 Hz (3 db cutoff) by a Grass Model 12 amplifier system. Activity over the right mastoid was actively recorded on a 16th channel to determine if there were any effects of the experimental variables on the mastoid recordings. No such effects were observed.

Continuous analog-to-digital conversion of the EEG and stimulus trigger codes was performed by a Data Translation 2801-A board and a Pentium-based computer at a sampling frequency of 200 Hz. Epochs were comprised of the 100 ms preceding and the 1180 ms following presentation of individual words in the sentences. Trials characterized by excessive eye movement or amplifier blocking were removed prior to

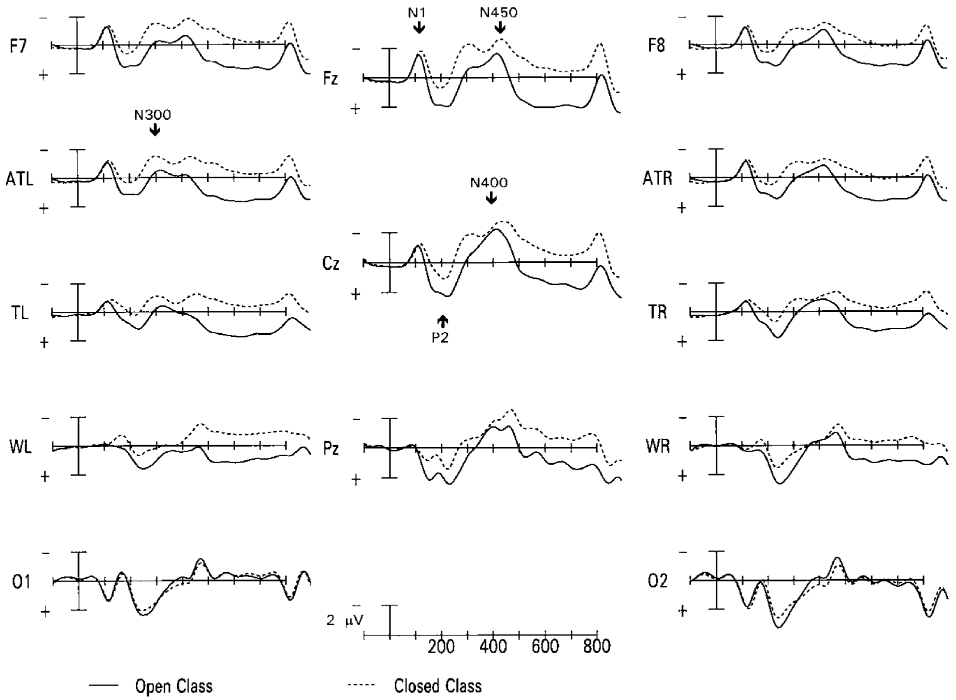

Fig. 1. ERPs recorded over 13 scalp sites to closed-class and open-class words. Each hashmark represents 100 ms. Positive voltage is plotted down.

averaging (approximately 7% of the trials were removed due to artifact for both the open- and the closed-class words).

ERP components of interest were quantified by computer as either mean voltage or peak (maximal) amplitude or latency within a window of activity, relative to the 100 ms of activity immediately following presentation of the word of interest. Window sizes are specified separately for each set of analyses. ERP measures were subjected to separate ANOVAs for midline and lateral sites. For midline sites, a three-way ANOVA model was used with a between-subjects variable of task and within-subject variables of word type and hemisphere. For lateral sites, a four-way ANOVA model was used with a between-subjects variable of task and the within-subject variables of word type, hemisphere, and electrode site. The Greenhouse–Geisser (1959) correction was applied to all repeated measures with greater than one degree of freedom in the numerator. In such cases, the corrected p value is reported. For analyses involving significant interactions between sentence types and electrode sites in the presence of a reliable main effect of sentence type, reported analyses are on data that have been normalized following the procedure described by McCarthy and Wood (1985). This normalization procedure was used because in certain cases spurious interactions can result if the experimental effects are of different overall amplitude. In order to minimize the number of reported analyses, analyses on normalized data are reported only when the effect is deemed to be theoretically important

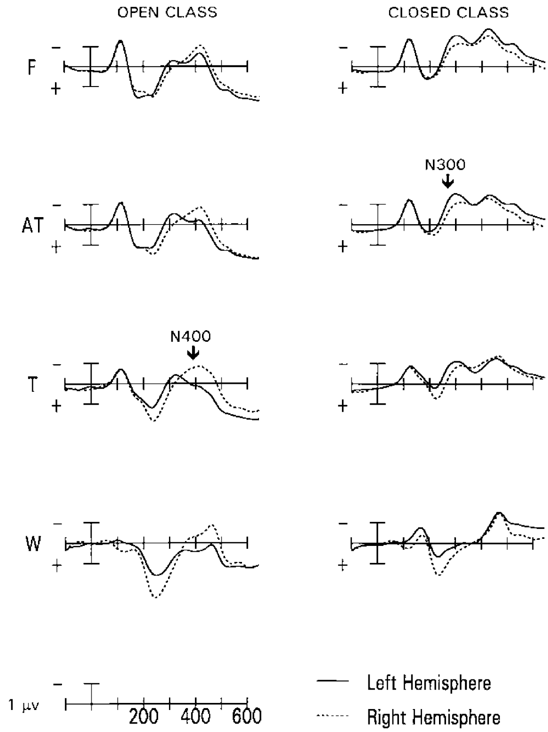

Fig. 2. ERPs to closed-class and open-class words, overplotting activity recorded over the left and right hemispheres.

(i.e. when the distributions of the responses to two conditions are directly compared). For analyses examining the effects of word length, regression analyses were also performed.

3. Results

3.1. Averages over word classes

Figs. 1 and 2 plot ERPs averaged as a function of word class (open versus closed), collapsing over the task variable. Inspection of these figures reveals that, consistent with many previous reports, ERPs to closed-class words were more negative-going than those to open-class words throughout the recording epoch, beginning at about 100 ms. In addition, a series of component fluctuations (similar to those previously reported; cf. Osterhout & Holcomb, 1992) were identifiable. A clear negative–positive complex was visible in the first 300 ms after stimulus presentation. The negative component (*N1*) peaked at about 120 ms. The positive component (*P2*) peaked at about 200 ms. The P2 was followed by a negative peak at about 300 ms (*N300*) that was most noticeable over anterior sites in the left hemisphere (F7, ATL). This component was elicited by both word classes but was

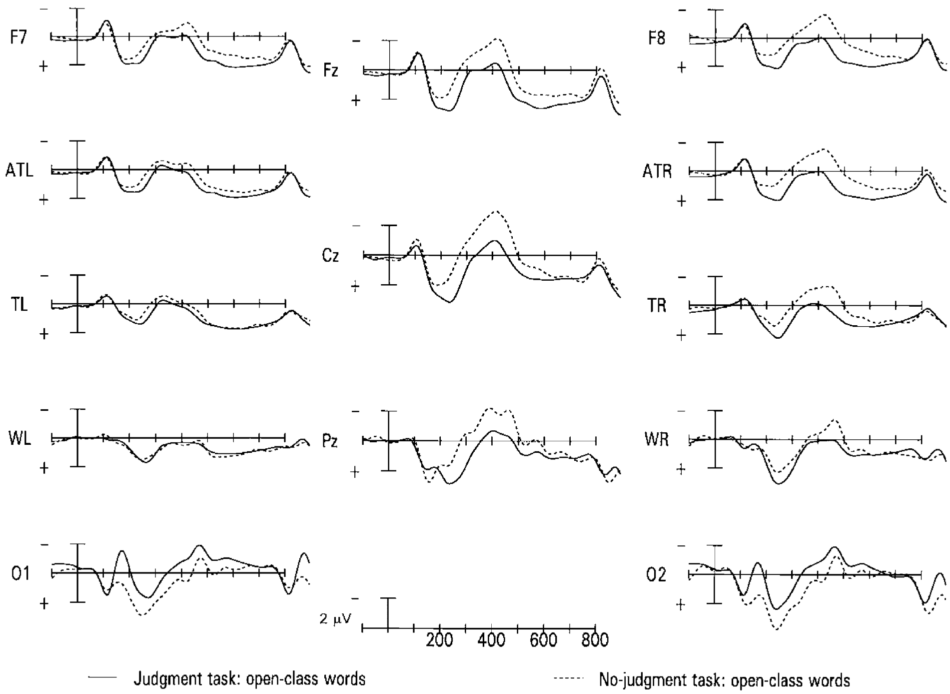

Fig. 3. ERPs to open-class words in the sentence-acceptability-judgment and passive reading conditions.

more negative in voltage for closed-class words. Open-class words also elicited a large negative peak at 400 ms (*N400*) that was largest over midline and right-hemisphere sites. Both word classes elicited another negative peak at about 450 ms (*N450*), probably reflecting word offset. Closed-class words elicited more negativity than open-class words from about 400–700 ms (*N400–700*).

These effects were quantified as mean amplitude within the following time windows: 150–280 ms (P2); 280–320 ms (N300); 350–500 ms (N400); and 500–700 ms (N400–700). P2 amplitude was larger for open- than for closed-class words (midline, $F(1,18) = 74.83$, $p < 0.0001$; lateral, $F(1,18) = 85.15$, $p < 0.0001$), particularly over anterior sites (word type × electrode site: midline, $F(2,36) = 4.11$, $p < 0.05$). The N300 component was more negative-going for the closed-class words than for the open-class words (midline, $F(1,18) = 28.02$, $p < 0.0001$), particularly over anterior sites (word type × electrode site: midline, $F(2,36) = 7.39$, $p < 0.01$; lateral, $F(4,72) = 9.11$, $p < 0.001$). In its temporal and morphological properties, this effect was highly similar to the N280 effect reported by Neville et al. However, unlike the Neville et al. report, this effect did not have a reliably larger amplitude over left anterior sites, relative to other sites (word class × hemisphere × electrode site: $F(4,72) = 1.35$, $p > 0.2$).

Throughout both the N400 and N400–700 epochs, ERPs to closed-class words were more negative than those to open-class words (N400: midline, $F(1,18) = 17.55$, $p < 0.001$; lateral, $F(1,18) = 21.36$, $p < 0.001$; N400–700: midline, $F(1,18) = 74.03$,

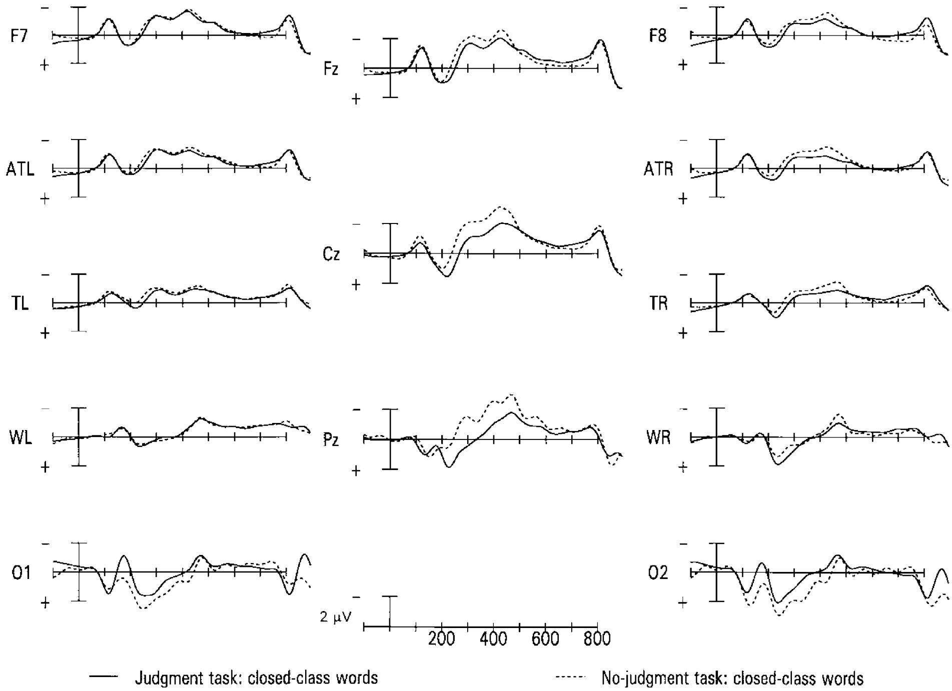

Fig. 4. ERPs to closed-class words in the sentence-acceptability-judgment and passive reading conditions.

$p < 0.0001$; lateral, $F(1,18) = 103.05$, $p < 0.00001$), and these differences were largest over anterior sites (word type × electrode site, N400: midline, $F(2,36) = 14.47, p < 0.001$; lateral, $F(4,72) = 25.78$, $p < 0.001$; N400–700: midline, $F(2, 36) = 16.56$, $p < 0.001$; lateral, $F(4,72) = 32.58$, $p < 0.0001$).

Figs. 3 and 4 plot ERPs to open- and closed-class words, respectively, as a function of task. For both word classes, the task manipulation primarily affected activity in the N400 window. Specifically, beginning at about 200 ms and continuing until about 500 ms, activity was more negative-going in the passive reading task condition than in the task requiring sentence acceptability judgments. These differences between task conditions were largest over anterior sites for open-class words and more evenly distributed in the anterior-to-posterior dimension for the closed-class words (task × word class × electrode site, 350–500 ms: midline, $F(2,36) = 4.79$, $p < 0.05$; lateral, $F(4,72) = 3.58$, $p < 0.05$). For both word classes, ERP differences between task conditions were larger over the right than the left hemisphere, although hemispheric asymmetries were more robust for open-class words (task × word class × hemisphere, 350–500 ms: $F(1,18) = 4.38$, $p = 0.05$). These effects can be reasonably attributed to task-related modulations of the N400 component; prior work has shown that N400 amplitude to words in context is largely a function of the semantic context in which the words occur (cf. Osterhout & Holcomb, 1995). One plausible explanation is that subjects who were making sentence-acceptability judgments after each sentence more completely (and more consistently) processed

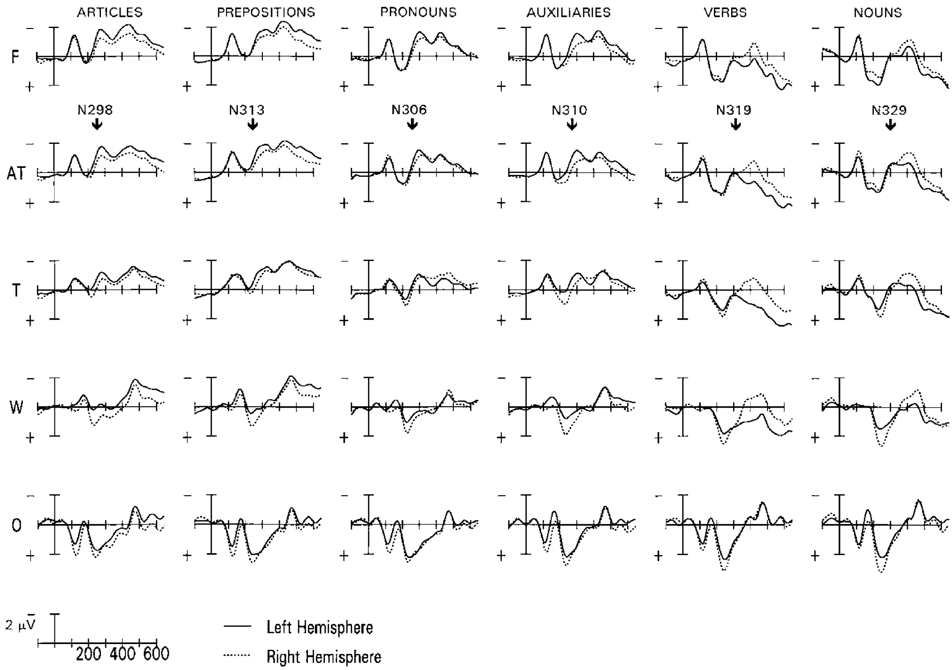

Fig. 5. ERPs averaged over grammatical categories, overplotting activity recorded over the left and right hemispheres.

the meaning of the discourse than did subjects given a passive reading task, thereby maximizing contextual effects on N400 amplitude. The slightly more robust effects for open-class words might be another manifestation of the frequently observed finding that open-class words elicit larger N400s (and show larger N400 effects) than do closed-class words. Regardless of the interpretation of these task effects, the important point for present purposes is that the task manipulation had similar (although not identical) effects on the ERPs to open- and closed-class words.

3.2. Averages over grammatical categories

Words were then subdivided on the basis of grammatical category. Grand-average waveforms, collapsing over the task factor, were computed for all grammatical categories that contained 90 or more exemplars within the text shown to participants. This criterion was chosen as it provided an adequate signal-to-noise ratio for examining the expected effects of interest. These subclasses included four types of closed-class words (articles, prepositions, pronouns, and auxiliary verbs) and two types of open-class words (nouns and verbs).

ERPs averaged over exemplars in each category are shown in Fig. 5. Consistent with the report of Osterhout et al. (1997), all categories elicited a negative-going component that peaked between 280 and 400 ms (for all grammatical categories, this negative component

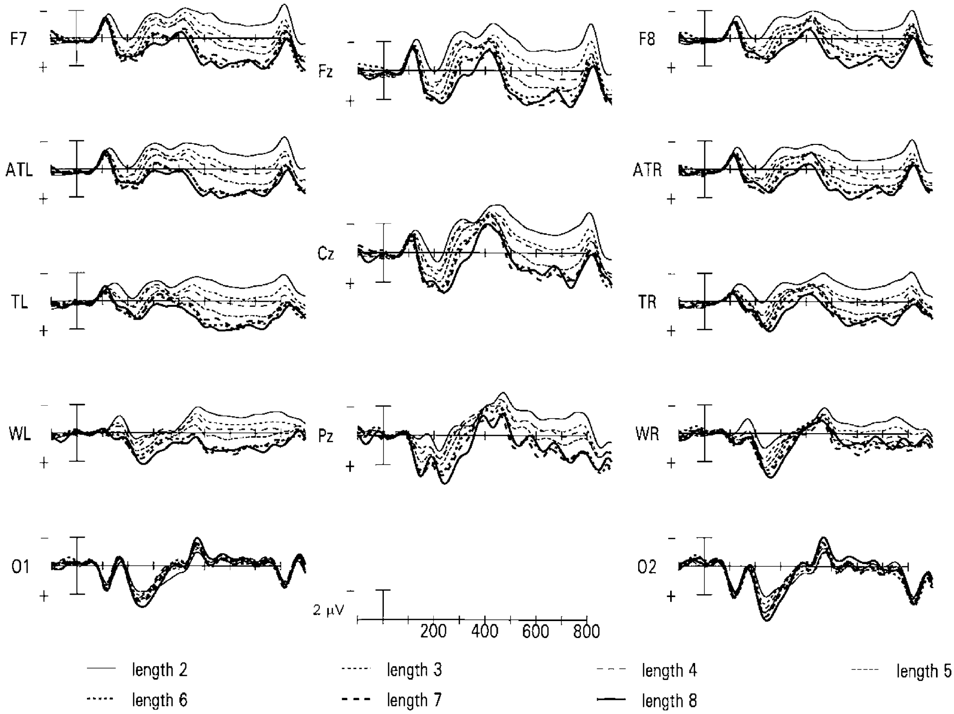

Fig. 6. ERPs averaged as a function of word length in letters, for words 2 to 8 letters in length.

was the first post-N1 negativity). The peak latency of this negativity varied across grammatical category. The peak was earliest-occurring for articles (N293), intermediate for prepositions (N313), pronouns (N306), and auxiliary verbs (N310), and latest-occurring for verbs and nouns (N319 and 328). The order of latencies across grammatical categories approximately correspond to the order of mean word frequencies and lengths in these categories (cf. Osterhout et al., 1997).

Statistical analyses were performed to determine whether the first post-N1 negative peak differed in latency across grammatical category. Peak latencies were determined by selecting the most negative value within a latency window of 250–350 for all sites except the occipital sites, where the relevant effects were not discernable. Analyses on peak latency found a robust effect of grammatical category on latency, $F(5,90) = 8.33$, $p < 0.001$.

3.3. Averages over word length

Averages were then formed separately for words of each length in letters, for lengths of 1 to 9 letters (Fig. 6). Inspection of Fig. 6 reveals that all words elicited a similar series of negative- and positive-going deflections (N1, P2, N300, N400, N450). Furthermore, beginning at about 100 ms and continuing until the end of the recording epoch, shorter words elicited more negative-going ERPs relative to longer words.

Table 1
Individual regression parameters (μV) and proportion of variance accounted for, for amplitude regressed onto word length

	Intercept	Slope	R^2	$p <$
P2 (150–300 ms)				
Midline	1.17	0.41	0.89	0.0001
Lateral	0.66	0.30	0.90	0.0001
N300 (280–320 ms)				
Midline	− 3.45	0.38	0.87	0.0001
Lateral	− 2.34	0.29	0.94	0.0001
N400 (350–500 ms)				
Midline	− 2.68	0.25	0.83	0.0006
Lateral	− 1.68	0.27	0.92	0.0001
N400–700 (500–700 ms)				
Midline	− 2.07	0.53	0.95	0.0001
Lateral	− 1.42	0.41	0.96	0.0001

Effects of length on ERP amplitude and latency were evaluated in two ways. First, measurements of peak or mean amplitude, and peak latency, within specific time windows were computed. ANOVAs were then performed on the data using length as a repeated measures factor. Second, in order to determine the strength of the relationship between word length and the waveform properties of amplitude and latency, measurements of amplitude and latency were regressed onto a model using word length as a single predictor.

For the ANOVA analyses involving amplitude measures, the following windows were chosen: peak amplitude between 150 and 300 ms (P2); mean amplitude between 280 and 320 ms (N300); mean amplitude between 350 and 500 ms (N400 region); and mean amplitude between 500 and 700 ms (N400–700 region). For midline sites, values were computed for Fz and Cz only. For lateral sites, values were computed over F7/F8, ATL/ATR, TL/TR, and WL/WR. The most posterior sites were excluded because visual inspection of the data suggested that robust differences between words of different lengths were not as evident at the most posterior sites. All of the analyses showed extremely robust effects of length on amplitude (150–300 ms: midline, $F(8,144) = 20.55$, $p < 0.0001$; lateral, $F(8,144) = 26.94$, $p < 0.0001$; 280–320 ms: midline, $F(8,144) = 14.60$, $p < 0.001$; lateral, $F(8,144) = 16.72$, $p < 0.001$; 350–500 ms: midline, $F(8,144) = 10.07$, $p < 0.001$, lateral, $F(8,144) = 15.66$, $p < 0.001$; 500–700 ms: midline, $F(8,144) = 25.59$, $p < 0.0001$; lateral, $F(8,144) = 28.35$, $p < 0.0001$). Between 150 and 500 ms, these differences in amplitude were largest over more anterior sites (length × electrode site: lateral, 150–300, $F(24, 432) = 2.58$, $p < 0.01$; 280–320, $F(24,432) = 3.02$, $p < 0.01$; 350–500, $F(24,432) = 4.16$, $p < 0.001$). Between 500 and 700 ms, differences between conditions were more evenly distributed in the anterior–posterior dimension (length × electrode site: $p > 0.2$).

Shifts in peak latency of the P2 and N300 components were quantified as the latency of the most positive-going point between 150 and 300 ms and the most negative-going point between 280 and 320 ms, respectively. Effects of word length on the latencies of both

Table 2
Individual regression parameters (ms) and proportion of variance accounted for, for peak latency regressed onto word length

	Intercept	Slope	R^2	$p <$
P2 (150–300 ms)				
Midline	200.67	2.7	0.69	0.005
Lateral	206.11	1.8	0.23	0.2
N300 (280–320 ms)				
Midline	303.49	1.1	0.51	0.03
Lateral	302.94	1.2	0.64	0.01

effects were robust (lateral sites: P2, $F(8,144) = 5.31$, $p < 0.01$; N300, $F(8,144) = 4.34$, $p < 0.01$).

Measurements of mean amplitude and latency obtained in the above analyses were then regressed onto word length. Throughout the recording epoch, and for both midline and lateral sites, amplitude was highly correlated with word length; length accounted for more than 80% of the variance in each analysis (Table 1). Variations in peak latency for the P2 and N300 were quite small. Nonetheless, these variations were robustly correlated with word length (Table 2).

One concern is that the effects of length have been confounded with the effects of horizontal eye movements. That is, longer words might induce more eye movements. Even though stringent criteria were used to eliminate eye artifact, it is conceivable that eye movements were sufficiently robust, and sufficiently systematically linked to word length, to introduce artifactual activity correlated with word length. Activity recorded via the face electrodes monitoring vertical and horizontal eye movements, plotted separately for words of each length, is shown in Fig. 7. Activity associated with both vertical and horizontal eye movements did show activity correlated with word length. However, two observations clearly indicate that this activity was cortical rather than ocular in nature. First, no polarity reversals were observed at electrode locations that would

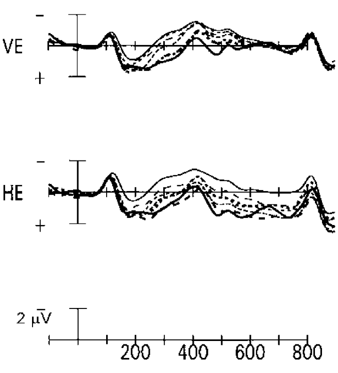

Fig. 7. Activity recorded over the vertical (VE) and horizontal (HE) eye channels, averaged as a function of word length for words 2 to 8 letters in length (see Fig. 6 for line descriptors).

be located on opposite sides of the dipole field, assuming an ocular source. Second, an ocular source would produce much larger potentials at the electrode positioned closest to it (i.e. the horizontal eye channel) than at the scalp electrodes positioned further away from it. However, the effects of word length were actually much more robust over the scalp sites Fz and Pz than over the horizontal eye channel. The fact that activity recorded over the horizontal eye channel most closely approximated that recorded over the closest scalp site (F8) also indicates a cortical, rather than an ocular, source.

4. Discussion

Neville et al. (1992) reported that open- and closed-class words elicit distinct components in the ERP, and attributed these differences to qualitative differences in the linguistic functions of these words. In the present study, we observed differences in the ERPs to open- and closed-class words similar to those reported by Neville et al. However, these differences were clearly a function of quantitative differences in word length: Variations in length accounted for almost all of the variance in amplitude and most of the variance in latency. Increases in word length were associated with increasingly positive-going activity throughout the recording epoch.

These results are seemingly inconsistent with the claim that word class has a robust influence on the ERPs elicited by words (Münte et al., 2001; Neville et al., 1992; Van Petten & Kutas, 1991). Our findings are more consistent with the results of two recent studies examining the effects of word length and frequency on ERPs to words (King & Kutas, 1998; Osterhout et al., 1997). These studies reported that the latency of an early negativity is correlated with word length and frequency. Here, we show that both the latency and the amplitude of successive components in the recording epoch are highly correlated with word length, beginning as early as 150 ms after word onset and continuing throughout the epoch. Because word length and frequency in our stimuli were highly correlated ($r = 0.99$), we cannot definitively determine which (or if both) of these variables underlies the robust correlations we report. However, there are reasons to believe that word length, rather than frequency, accounts best for the results reported here. The correlation between word length and frequency is of an inverse nature, such that longer words tend to have lower frequencies and shorter words tend to have higher frequencies. As observed in Neville et al. (1992) and other studies, the expected effect for word frequency, if any, is one in which lower frequency words elicit more *negative* going wave forms during the N400 epoch. In the present study, though, we observed progressively more *positive* going wave forms during this epoch as a function of length—which is the opposite of what one would expect had the effects at hand been driven predominantly by word frequency. Furthermore, the effects of word frequency on the ERP waveform (and on most other measures as well; e.g. lexical decision times) are greatly reduced when words are presented in sentence or discourse contexts (as in the present study), compared to when they are presented in single-word processing tasks (Van Petten & Kutas, 1991).

We should reiterate that when we grouped our data by word class (open versus closed) only, our results appeared to closely match those reported by both Neville et al. (1992) and Brown et al. (1999).[2] These researchers evaluated frequency and length effects, but concluded that these variables could not account for the crucial differences in ERPs that they associated with more abstract, linguistic properties of words. Our belief is that the discrepancy is largely a function of the various methods used in these studies for quantifying the effects of word length and frequency. In the present study, the effects of length on the ERP waveform were directly assessed by creating separate averaged waveforms for words of each length, within a range of 1 to 9 letters. The relationship between word length and changes in the ERP waveform was assessed by employing both an ANOVA model and a regression model. By contrast, Neville et al. and Brown et al. created a small number of length or frequency bins by averaging over a small range of lengths or frequencies and examined mean differences using an ANOVA model. Specifically, Neville et al. formed two averages corresponding to high- and low-frequency or to short and long open-class words. Brown et al. compared ERPs to words of a few selected lengths (e.g. lengths of 3 versus 4 letters, or 5 versus 6 letters) that were of similar frequency, or to words of different frequencies, but similar lengths. Neither team of researchers reported any compelling effects of word length or frequency.

The most definitive approach to determining the strength of the relationship between word length or word frequency and variations in the brain's electrical activity, in our view, is to perform regression analyses on as wide a sampling of word frequencies or lengths as is practical. Regression analyses are critical because they provide direct estimates of the strength of the relationship. Severely restricting the range of these variables, as was done by both Neville et al. and Brown et al., might obscure the nature of these relationships. It is conceivable that a re-analysis of the Neville et al. and Brown et al. data using the statistical approach taken in the present study would produce a consistent outcome in all three studies.

To the best of our knowledge, these effects of word length on ERPs have not been previously reported. The existence of such robust effects introduces a methodological caution. It appears that word stimuli must be closely matched in length across conditions in order to avoid confounding the effects of interest with the effects of word length.

In summary, our findings suggest that the open- and closed-class distinction does not obviously and robustly manifest itself in the standard time-domain analysis of scalp-recorded event-related potentials. However, our findings in no way imply that these (and other) lexical category distinctions are not represented in the brain. Indeed, it would be difficult to account for speakers' linguistic competence if at least some of the linguistically motivated lexical distinctions were not somehow represented in the brain. It could very well be that the electrophysiological concomitants of these distinctions will become visible only with application of other techniques (e.g. coherence and power analyses) for analyzing the brain's response to words.

[2] Neville et al. (1992) reported that the N280 was larger over the left than the right hemisphere, whereas we did not observe a robust lateralization for this effect. This difference might reflect, in part, the use of different reference electrodes in the two studies. Neville et al. used a linked mastoid reference, whereas a left mastoid reference was used in the present study.

Acknowledgements

This research was supported by research grants R29 DC 01947 and R01 DC01947 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, awarded to LO.

References

Bradley, D. C., & Garrett, M. F. (1983). Hemispheric differences in the recognition of closed- and open-class words. *Neuropsychologia*, *21*, 155–9.

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: open- and closed-class words. *Journal of Cognitive Neuroscience*, *11*, 261–81.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language*, *3*, 572–82.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: an introduction. In M. D. Rugg & M. G. H. Coles, *Electrophysiology of mind: event-related brain potentials and cognition*, Oxford: Oxford University Press.

Damasio, H., & Damasio, A. R. (1989). *Lesion analysis in neuropsychology*, New York: Oxford University Press.

Garnsey, S. M. (1985). *Function words and content words: reaction time and evoked potential measures of word recognition*. (Cognitive Science Tech. Rep. No. URCS-29). Rochester, NY: University of Rochester.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth, *Language production, vol. I, speech and talk*New York: Academic.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Heilman, K. M., & Scholes, R. J. (1976). The nature of comprehension errors in Broca's conduction, and Wernicke's aphasics. *Cortex*, *12*, 258–65.

Jasper, H. H. (1958). Report to the committee on methods of clinical examination in electroencaphalography: the ten–twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–5.

Johnson Jr, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*, 90–97.

Kenan, L. R. (1986). *New American profiles*, New York: Harcourt Brace Jovanovitch.

King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, *244*, 61–64.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*, 539–50.

Kutas, M., & VanPetten, C. (1994). Psycholinguistics electrified: event-related brain potentials investigations. In M. A. Gernsbacher, *Handbook of psycholinguistics* (pp. 83–143). San Diego: Academic Press.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, Sup.*, *39*, 325–30.

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*, 203–8.

Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia*, *39*, 91–102.

Neville, H. J., Mills, D. J., & Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*, 244–58.

Nobre, A. C., & McCarthy, G. (1994). Language-related ERPs: scalp distributions and modulation by word type and semantic priming. *Journal of Cognitive Neuroscience*, *6*, 233–55.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L., & Holcomb, P. (1995). Event-related brain potentials and language comprehension. In M. D. Rugg

& M. C. H. Coles, *Electrophysiology of mind: event-related brain potentials and cognition*, Oxford: Oxford University Press.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: word length and frequency predict the latency of an early negativity. *Biological Psychology*, *46*, 143–68.

Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology*, *94*, 357–70.

Saffran, E. M., Schwartz, M. F., & Marin, O. S. M. (1980). The word order problem in agrammatism. II: production. *Brain and Language*, *10*, 249–62.

Schwartz, M., Saffran, E., & Marin, O. (1980). The word-order problem in agrammatism. I. comprehension. *Brain and Language*, *10*, 263–80.

ter Keurs, M., Brown, C. M., Hagoort, P., & Stegeman, D. F. (1999). Electrophysiological manifestations of open- and closed-class words in patients with agrammatic comprehension: an event-related potential study. *Brain*, *122*, 839–54.

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