JOURNAL OF MEMORY AND LANGUAGE **37,** 217–239 (1997)
ARTICLE NO. ML972514

Sentential and Acoustic Factors in the Recognition of Open- and Closed-Class Words

Daniel T. Herron and Elizabeth A. Bates

Department of Cognitive Science, University of California, San Diego

Closed-class words are highly frequent yet relatively difficult to perceive; although this ought to impair communication, we communicate easily under normal conditions. Modular and interactive architectures offer differing explanations of this paradox, with different assumptions about how the acoustic and grammatical properties of those words are combined. The interaction of these factors was investigated by having subjects listen for and repeat open- and closed-class homophones (spoken by a male) that were spliced into three female-voice sentences: (a) the same sentence, (b) a neutral sentence, and (c) the ''swapped'' sentence (e.g., open- target in a closed-class context). Results show that: (a) under neutral conditions, it is harder to identify closed- than open-class tokens; but (b) they differ little in their original contexts; (c) open-class tokens are very easy to identify in a closed-class context; (d) recognizability of closed-class tokens in the swapped context was generally poor; and (e) these interactions are influenced by sentence prosody but not by target length. It is argued that these results indicate a relatively early interaction between perceptual and contextual processing.

The distinction between open-class (or content) and closed-class (or function) words is defined, somewhat arbitrarily, on the productive nature of each class; namely, the open class adds new members easily, while the closed class does not. The words differ between classes in other ways, however: closed-class words are used primarily to express grammatical and semantic relations, are generally very high in frequency and low in semantic content, and are acoustically less salient (making them difficult to perceive out of context, as shown long ago by Pollack & Pickett, 1964). When in their normal context, however, closed-class words are generally perceived and processed with a high degree of accuracy, as should be obvious given the success of linguistic communication.

Support for this research was provided by a grant to Elizabeth Bates, ''Crosslinguistic Studies in Aphasia'' (NIH/NIDCD Grant 2-RO1-DC00216). We thank Jeff Elman, Mary Hare, Arturo Hernandez, Javier Movellan, and Lynn Snyder for helpful discussions.

Address correspondence to Daniel Herron, Center for Research in Language 0526, University of California at San Diego, La Jolla, CA 92093-0526. E-mail: herron@cogsci.ucsd.edu.

Various theories attempt to account for this paradox; all theories must, however, explain how grammatical and acoustic information is used and combined in the processing of those words. The four experiments described in this paper are designed to explore the similarities and differences in the processing of these words. (Note that most of this paper concerns open- and closed-class words specifically in English; although some findings may translate to other languages, no such specific claims should be or are implied.) Specifically, we investigated the contribution of acoustic and contextual cues to the recognition of open- and closed-class homophones (words that are highly similar in form but markedly different in function) when they are placed in the appropriate context. Before these studies are presented, we will briefly review evidence that open- and closed-class words are indeed processed differently, together with theories of how, why, and when this is done.

PSYCHOLINGUISTIC EVIDENCE

Differences in the processing of open- and closed-class words have been postulated based on a variety of evidence. Differing probabili-

0749-596X/97 $25.00

ties of occurrence in speech errors (e.g., one occasionally encounters errors such as ''the pot of gold'' → ''the gold of pot'' but never ''the pot of gold'' → ''of pot the gold'') led to the claim that open- and closed-class words are represented, planned, and processed at different levels of speech production (Garrett, 1975). Differences in recognition have also been found; in a series of lexical decision tasks, Bradley (1978) showed that closed-class words exhibit neither the same frequency effects as open-class words (for which reaction time [RT] is inversely related to frequency) nor the same interference effects when placed at the beginning of nonwords (e.g., ''*such*erty'' was rejected as a nonword more quickly than was ''*word*erty'').

Bradley (1978) argued that because a sentence-processing mechanism can plausibly entertain various phrasal analyses based on the same input (given the high degree of ambiguity in natural languages), there must be some ''islands of certainty'' on which a listener can rely, and from which he or she can then parse the rest of the sentence. For computational efficiency, she argued, the language-processing device uses special access procedures for these islands of certainty (which are, of course, the closed-class words); i.e., the system uses different mechanisms for open- and closed-class forms. It must be noted, however, that there is considerable dissent in the literature over the findings on which this claim is based (Garrett, 1975; Matthei & Kean, 1989; Petocz & Oliphant, 1988).

The strength of this claim was further magnified by the finding that Broca's aphasics tested in the same manner as the normal controls failed to show the dissociation between open- and closed-class words (Bradley, 1978; see also Church, 1987). Specifically, the aphasics seemed to treat closed-class words in the same manner as they (and normals) treated open-class words—recognition of closed-class items varied directly with frequency, and both classes of words produced increased latencies when embedded in nonwords—as if the special closed-class pathway was damaged or otherwise unavailable. Lateralized presentation of visual stimuli has also brought out these differences; the left hemisphere has been claimed to show the normal differences in processing, while the right hemisphere, like the aphasic patients, maintains only one word recognition device (Bradley & Garrett, 1983; Shapiro & Jensen, 1986; although Chiarello & Nuding, 1987, found no evidence for such a hemispheric difference.)

PROCESSING MODELS

Various models can explain the differences in processing observed between open- and closed-class words. The three described below may usefully be considered to range from ''maximally modular'' (meaning that only bottom-up information is available prior to lexical access) to ''maximally interactive'' (meaning that all relevant information sources are potentially used in achieving lexical access), although they are not strictly oriented along such an axis. They vary, as well, in whether or not they claim that word-class distinctions cause processing distinctions prior to lexical access. Of course, neither a strict modular nor a strict top-down system is implausible a priori, and thus the issue can be decided only based on empirical evidence (Swinney, 1982; Tanenhaus & Lucas, 1987).

(a) Separate Processors

If separate processors are devoted to open- and closed-class words, some sort of information must be able to determine which processor should be used. This information could be bottom-up (e.g., the physical differences in the realization of the word-classes) or it could be top-down contextual clues (e.g., the knowledge that determiners are often followed by nouns). Two caveats must be addressed, however. First, contextual information should not be available to a strictly modular architecture until lexical access is complete. Second, because neither source of information can predict with perfect accuracy the form class of a word, any theory that proposes separate processing streams must be prepared to explain what happens when a word is handled by the wrong processor.

Some support for the claim of separate processors comes from the finding that aphasic patients show faster responses to open-class than closed-class words (Swinney, Zurif, & Cutler, 1980), while normals show no such difference in RT (Cutler & Foss, 1977), as if the aphasics have lost the ability to use the ‘‘special access and retrieval process’’ for closed-class items. By this account, normal processing uses two separate but essentially equal processors; thus testing of normals reveals *no* difference in RT as a function of word class. Metrical stress information (i.e., whether the syllables are strong or weak) is predictive of form class, and the normal processor can thus decide which processor to use based on the phonological characteristics of a word (Swinney et al., 1980). Of course, any mismatch between that cue and the word’s actual form class should engage the wrong processor; as Swinney et al. (1980) claim, closed-class words that happen to carry stress (e.g., ‘‘you DO love me, don’t you?’’) will receive the special attention that is normally given only to open-class words.

Listeners have, in fact, been found to be highly sensitive to the robust and consistent differences in the realization of open- and closed-class words, which are reflected in the sentence-level prosodic structure, and they seem to use these prosodic cues in order to locate word boundaries (Cutler, 1993). Cutler presented indirect evidence that listeners use a metrical segmentation heuristic: assume that any strong syllable (any syllable containing a full vowel) marks the beginning of an open-class word. The motivation for making such determinations is not clear, however; one can argue that faster recognition is more important for either word class. (Grosjean & Gee, 1987, in a related proposal, argued that processing varies along a continuum based on the strength of the syllables, so that even open-class words can in some cases be de-stressed.)

Both Cutler (1993) and Swinney et al. (1980) claimed that processing of stressed closed-class words is not impaired in normal listeners. In fact, Cutler dismissed the notion that there is *any* difference in the difficulty of processing open- versus closed-class items, although she did accept that closed-class words are generally harder to perceive (simply because they often have reduced vowels):

> There is no evidence that processing of closed-class words presents difficulty to the listener, despite their typical realization in phonologically weak form. . . . That is, closed-class words may in practice be hard to perceive, but in principle they are not otherwise hard to process. . . . In fact it would be highly surprising if they *were* hard to process; recall that closed-class words make up more than 50% of all word tokens occurring in typical speech samples (Cutler & Carter, 1987). If this high a proportion of all words we hear were to cause processing difficulty, then at the very least one might feel that our processing mechanism was not functioning optimally. (p. 117)

Thus because closed-class words are less perceptible, the processor responsible for them must be in some sense superior to the open-class processor, so that processing of the two classes of words presents roughly the same difficulty to the normal language processor. Cutler therefore predicted that unstressed open-class words should cause difficulty, as they force the less efficient processor to work with the less perceptible input.

(b) Parallel Permeable Modules

Friederici (1985; see also Friederici, 1982) argued that the lexical and nonlexical levels of processing operate simultaneously but independently. This claim was based on a word-monitoring study that found that varying the functional role of closed-class words affected their processing such that when closed-class words were more lexical in nature, their processing seemed more similar to that of open-class words. The distinction between lexical and nonlexical is not quite the same as that between open- and closed-class words—closed-class words can carry some lexical information—yet the two overlap highly. Thus while Friederici argued that phonological information plays a part in form-class distinctions, she abandoned a strictly modular model and allowed that higher-level contextual information can penetrate and affect the process of lexical access. She maintained a weaker form of autonomy, however, in claiming that higher-level information does not cross between the lexical and nonlexical levels (i.e., recognition of closed-class words will generally be facilitated only by preceding syntactic information, and recognition of open-class words will generally be facilitated only by preceding semantic context).

A similar account of processing holds that processing of open-class words is a purely bottom-up process, while processing of closed-class words is sensitive to contextual information (this is computationally feasible because of the closed class's small size). Such a *restricted modularity hypothesis* (Shillcock & Bard, 1993) was supported by evidence that closed-class words (in a sentence context) that were homophonic with open-class words (such as *would* or *can*) did not prime the open-class meaning (e.g., *would* did not prime *timber*), even though stress differences between the open- and the closed-class words were controlled. The lack of closed-to-open priming was claimed to result from contextual permeability within the closed class. The benefit in allowing this limited degree of interaction with contextual information is that it can assist a ''prosodic sorter'' (e.g., Cutler, 1990; Cutler & Foss, 1977; Swinney et al., 1980), which was estimated by Shillcock and Bard to otherwise make errors on one quarter of all words. The Shillcock and Bard hypothesis is in agreement with Friederici's (1985) aforementioned hypothesis, because in neither case does syntactic information affect the recognition of open-class words. (Friederici's position, however, was not that open-class words per se are sensitive to semantic context, but that processing at the semantic level—which is generally equivalent to processing of open-class items—is sensitive only to semantic context.)

(c) The Interactive Alternative

One alternative to the preceding theories is that language processing does not divide open- and closed-class words (or any two classes) at any distinct point in time. For example, the competition model of Bates and MacWhinney (1989) is a framework for the study of cross-linguistic facts, designed to account for the rapid integration of all available cues. Instead of postulating separate processors for various classes of words, phonological and contextual information can be used simultaneously. All cues that might help in the processing are considered at the same time (or as soon as they are available) and given weight in accord with their reliability and validity in the listener's native language. The mapping between form and function is thus viewed as a kind of constraint satisfaction that takes place without intermediate steps or processing divisions. Differences between classes of words (nouns and verbs or open- and closed-class) lie in the ''access properties'' of individual items (for example, their frequency, salience, imageability, uniqueness, or demands they place on memory) and not on the processors that control lexical access per se (Bates & Wulfeck, 1989; see also MacDonald, Pearlmutter, & Seidenberg, 1994). Although an interactive theory such as this postulates very different mechanisms, it would draw on the same sources of information that have been considered above—the major differences lie in how and when those cues are combined and considered.

One cannot choose among these theories without knowing more about their exact architecture and more about the time course of processing (e.g., how long lexical access is presumed to take.) One would also need very sensitive techniques to address this issue, techniques that are perhaps not yet available. Nevertheless, a study of how various cues and constraints affect processing—even given a relatively crude measure of processing—might prove informative. Thus, in this paper we present a series of RT studies focused on the recognition of open- and closed-class words. The major goal was to understand the interaction between sentential and acoustic information in processing of those words.

In Experiment 1, homophonic open-class/closed-class pairs (e.g., *dew/do* or *mite/might*) were used as targets in an auditory cued shadowing task, so that the sentential (contextual) and acoustic factors could be manipulated orthogonally. Experiments 2 and 4 (using the same task) and Experiment 3 (using a variation on the task) were designed to reveal what, specifically, were the relevant aspects of the contextual and acoustic factors. The cued shadowing task, as used here, involves presenting subjects with a sentence read by a speaker, with one word in the sentence (the target) read by a speaker of the opposite sex. This technique has been used successfully by Liu, Bates, Powell, and Wulfeck (in press) in a study of word frequency and semantic context on lexical access. Liu et al. argued that the technique has several advantages over others: (a) it can be done in a purely auditory setting (which is arguably more natural, because language evolved as speech), (b) the response is likely to be less clouded by ‘‘decision’’ components than in some other tasks (such as lexical decision), and (c) subjects do not need to know the identity of the target word in advance (which is again more natural and also less likely to introduce confounds). The cued shadowing task is presumed to provide a valid measure of the difficulty of receptive lexical access, although access to the motor code is relevant as well.

GENERAL METHOD

The experiments described below were highly similar in method. In order to emphasize this similarity, the general method will be presented first, and only the minor deviations from this method will be detailed with each experiment.

Stimuli

A set of 126 sentences[1] was generated for these experiments. Each sentence consisted of two clauses, with the target word appearing in the second clause (never as the initial or final word). Seventy-two of the target words were homophones (36 open-class and 36 closed-class words, listed in the Appendix); these served as the experimental items. Note that the use of open- and closed-class words as targets (especially when limited to homophonic pairs) made it impossible to control for frequency. The open-class targets had an average written frequency of 134 occurrences per million, while the closed-class had an average of 3191 (Kucera & Francis, 1967).

The remaining 54 targets were filler items; 38 of these consisted of 20 open-class and 18 closed-class words that did not appear as experimental targets. Sixteen of those 38 fillers (half open- and half closed-class) were repeated (used as targets in two different sentences), to produce a total of 26 closed-class and 28 open-class filler sentences. The use of fillers that appeared one or two times was designed to distract subjects from noticing that the experimental targets were eventually followed by their similar-sounding homophones. An additional set of eight neutral sentences (e.g., ‘‘The next sentence contains a word to repeat; please say . . . at this time’’) was also created.

All sentences were recorded by both a male and a female speaker, in a sound-proof chamber, using a high-quality microphone and a Sony digital audio tape recorder. The sentences were then sampled into a Macintosh computer (16 bits per sample and 22 kHz sampling rate). The root mean-squared (RMS) amplitude of the two recordings (male and female) was then normalized to ensure that the target words (male voice) would be roughly as loud as the sentences into which they were placed (female voice).

The degree of cloze-probability of each sentence (other than the neutral contexts) was determined in a pilot study. Twenty-one college-aged native speakers of American English, drawn from the subject pools of the Psychology and Cognitive Science programs, listened to each sentence (as read by the female) up to (but not including) the target word. The order of the items in this list was randomized for each subject; however, to control for any possible repetition effects, the sentences corresponding to the homophonic pairs were bal-

[1] The complete list of items used in this experiment is available electronically on the internet in the ‘‘publications’’ section of ⟨http://crl.ucsd.edu/⟩.

anced between the first and the second half. Subjects listened to the sentences and were instructed to guess verbally what the single most likely next word was. Subjects were given a window of 1500 ms in which to respond. Overall, subjects guessed the target word in the closed-class contexts 26.9% of the time and 27.4% in the open-class contexts. An ANOVA by subject (*F*1) and by item (*F*2) showed no significant difference between the open- and closed-class contexts in the number of target words guessed (*F*'s < 1). Thus, the open- and closed-class contexts seem to constrain the target words to a similar degree.[2]

To create the actual stimuli, the target words were physically isolated in both the male and the female versions of each sentence. The male version of the target word was then spliced into the female sentence in place of the word in target position. Because the sentences were read in a fairly normal manner (i.e., the speakers were instructed to speak clearly, but not slowly), it was impossible to perfectly isolate one word from the next. Nevertheless, considerable time was spent to make sure that there was as little distortion as possible on both sides of the operation. The actual quality of the stimuli was high; aside from the intended distortions caused by the splicing, the sentences sounded quite normal.

Each of the 72 experimental target words was also spliced into the sentence from which its homophonic mate was taken (again, splicing the male version into the female). For example, *dew* was placed into the sentence using *dew* as a target and into the sentence using *do* as a target; conversely, *do* was also spliced into both of those two contexts. Each experimental target word was also placed into one of the neutral-context sentences. Thus each of the 72 experimental target words was placed into three sentences: its original sentence, a ''swapped'' sentence, and a neutral sentence. Filler targets were placed only in their original sentences. This produced a total of 216 experimental sentences and 54 filler sentences.

The set of 216 experimental sentences was divided into three lists, each with 72 sentences. Each list had an equal number of open- and closed-class words in each of the three contexts (i.e., 12 sentences with an open-class target in its original sentence, 12 with different open-class targets in neutral contexts, and 12 with the remaining open-class targets in the swapped contexts; the closed-class items were similarly distributed.) In this way, exactly one physical recording of each of the 72 experimental target words appeared in each list; the two elements of each homophonic pair were placed in different context sentences (thus if *dew* appeared in a closed-class context, *do* appeared in either the neutral or the open-class context in that same list.) No subject received any context sentence more than once (except for the neutral contexts); every subject heard each target word exactly once. Furthermore, the lists were divided into halves, such that the open-class and closed-class tokens of each homophonic pair appeared in opposite halves (e.g., if *dew* appeared in the first half of the list, *do* appeared in the second.) Each list thus contained 72 experimental items; the 54 distractor items were then randomly distributed in each list, 27 in each half of the list. Subjects were assigned to one list, the sentences in which were randomized for each subject according to the above constraints.

The male and female speaker each recorded an additional 40 words, none of which appeared as target items in any sentence. These words were used in the baseline section of the experiment. In addition, 10 sentences identical in form to the experimental and distractor sen-

[2] Comparison of the number of incorrect responses sharing the same first letter with the actual target word allowed a crude orthographic measure of coarticulatory effects; if some trace of, for example, *do* or *dew* were still present in the stimuli, subjects might have been biased to respond with words beginning with ''d''. A relation between this bias and the word class might have formed a potentially troubling confound. Closed-class contexts were more likely (21.3%) than open-class contexts (16.5%) to lead to the correct first letter in incorrect responses, but the difference was not reliable ($F1(1,20) = 3.65$, $p = .070$; $F2(1,70) < 1$). We have confidence that the amount of coarticulatory information present prior to the target is not confounded with word class.

tences (with different target words) were prepared for a practice section.

Subjects

A separate group of 21 students drawn from the Psychology and Cognitive Science subject pools was used for each experiment.

Procedure

Subjects were tested in a quiet room using Macintosh computers and the PsyScope experimental package (Cohen, MacWhinney, Flatt, & Provost, 1993). The stimuli were presented either using speakers on the table in front of the subject or else using a pair of high-quality headphones. The voice key was triggered by a microphone that was either placed on the table or (when using headphones) attached to the headphones. The change of equipment was made (in the later experiments) to the combined headphones and microphone in order to more cleanly present the stimuli and capture the responses.[3]

After informed consent was obtained, the subjects in all experiments were presented with a baseline section in which pairs of words were presented and the second word was to be repeated; this section was used to allow the subjects to become comfortable with the equipment and for the experimenter to adjust the voice key. The baseline was followed by a practice session, which consisted of 10 trials using the practice stimuli presented in the same manner as the experimental trials in that experiment. Subjects were allowed to repeat the practice trials if they desired. Subjects were given rest breaks periodically during all experiments and were allowed to take additional breaks whenever they desired.

Each trial consisted of the presentation of one sentence read by a woman, which was composed of a context clause (e.g., ‘‘The doctor thought the patient had an infection’’) followed immediately by the target-bearing clause (e.g., ‘‘she couldn’t believe that a sting from a . . . could hurt that much’’). Subjects were instructed to ‘‘repeat as quickly as possible the word that was read by the man.’’ Subjects were asked to guess if they felt that they might know what word it was, but not to guess if they could not say a real word. A voice-keyed response time was collected, relative to the onset of the target word; the duration of the response window was 1500 ms. The responses of the subjects were tape recorded for later analysis by the experimenters.

Each trial was scored as either correct, omitted (the subject did not respond within the allowed time), or substituted (the subject responded with an incorrect word.) Responses were judged as correct by listening to the taped session; there was, of course, some difficulty owing to the use of homophonic targets. In most cases correct responses sounded much like canonical (out of context) pronunciations (in which case the open- and closed-class words are really essentially homophonic); it is thus possible, although doubtful, that subjects might have been responding with *dew* when the actual target was *do.* In a few cases, subjects appeared to imitate the *sound* of some of the more difficult closed-class words, even though they had been instructed to guess only ‘‘real words.’’ Such responses were generally dealt with conservatively and not scored as correct.

EXPERIMENT 1

Method

Subjects each received one of the three lists of 126 items described under General Methods. The procedure described under General Methods was followed, with auditory presentation of all stimuli.

[3] To confirm that the change in equipment had no effect, Experiment 1 was run again with 21 new subjects using the combined apparatus. The only significant difference caused by the change in equipment was a significantly higher overall accuracy with the headphones (82.3%) than with the speakers (78.4%) ($F1(1,40) = 5.30$, $p < .05$; $F2(1,35) = 11.27$, $p < .05$). There was no significant interaction with any other factors, however, and no significant effect on RTs. Thus we are confident that this change in equipment resulted in cleaner data without changing the overall pattern of results.

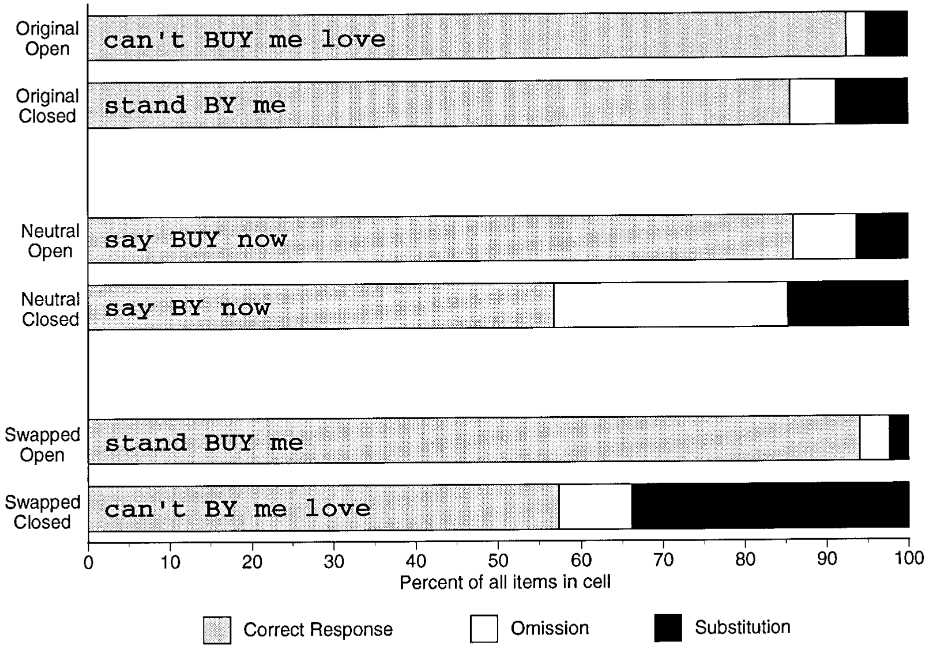

FIG. 1. Percentage of trials in Experiment 1 that were scored as correct, omitted, or substituted, as a function of target word class and sentence context. The example fragments are not representative of the actual stimuli.

Results

Two measures of the performance of the subjects (accuracy and RTs) were calculated. Overall ANOVAs were performed for both of these dependent variables (for the experimental items), to examine the main effects of and interaction between context (either original, neutral, or swapped) and target (either open or closed). Two additional sets of post-hoc tests were made for both dependent variables (and shall be repeated for each experiment). First, the contextual facilitation of responses relative to neutral was tested (i.e., the increase in accuracy or decrease in RT that results from being in a non-neutral context); these scores were calculated for both the open- and closed-class targets as (1) original-context minus neutral-context and (2) swapped-context minus neutral-context. Second, the difference between the open- and the closed-class targets within each level of context was also tested (i.e., open minus closed, for each of the three contexts).

Mean values reported here are averages across subjects (corresponding to $F1$); in the case of the RTs, the averages across items (corresponding to $F2$) differed from the averages across subjects in those cells with fewer correct responses, because the RTs are based only on the correct responses. Furthermore, for the ANOVAs by items, it was in some cases necessary to substitute in the cell mean for particular items that had no correct responses (and thus no valid RTs). Because of the uneven distribution of valid RTs across conditions, and because the RT data tended to mirror the accuracy data, the RT data will not be discussed in detail. The relevant figures and tables of RT data are, however, provided for the interested reader.

As seen in Fig. 1, closed-class items were significantly harder to identify; overall, subjects correctly identified 91% of the open-class targets and only 67% of the closed-class targets ($F1(1,20) = 347.43$, $p < .05$; $F2(1,35) = 27.09$ $p < .05$). There was a significant

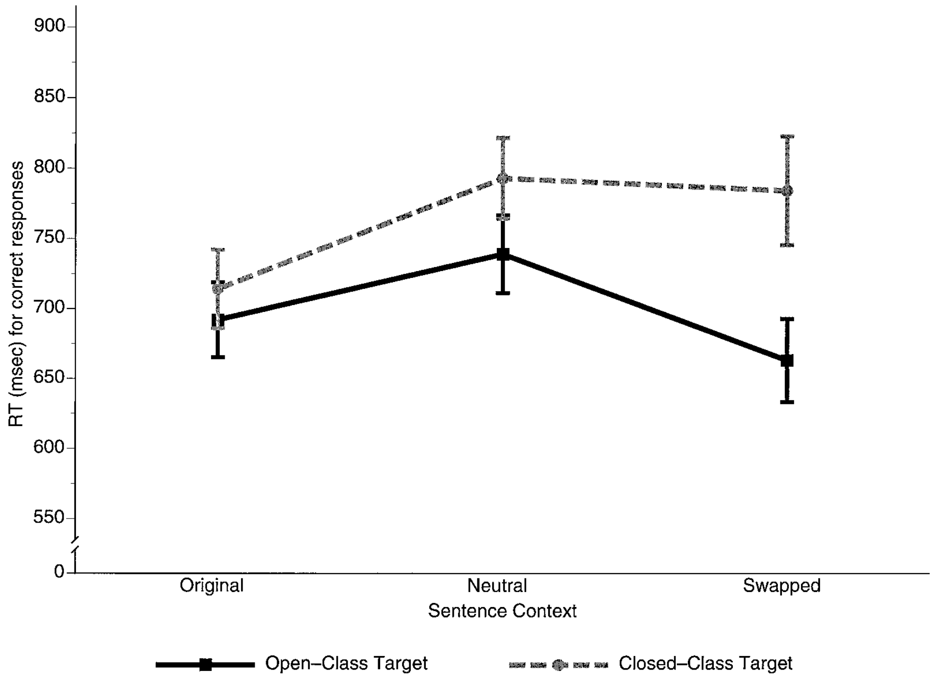

FIG. 2. Mean RT across subjects to correct responses in Experiment 1, as a function of target word class and sentence context. The error bars represent the standard error across subjects.

main effect of context; the original contexts were easiest (89%), followed by the swapped contexts (76%), followed by the neutral contexts (71%) ($F1(2,40) = 22.02$, $p < .05$; $F2(2,70) = 20.64$, $p < .05$).

There is also, however, a significant Context × Target interaction ($F1(2,40) = 14.17$, $p < .05$; $F2(2,70) = 16.31$, $p < .05$). The interaction reflects the greater influence of sentence context on the closed-class targets: Closed-class targets ranged in accuracy from 57% out of context to 86% in their original context, while the open-class targets never fell below 86% and showed no decrease in performance when placed in the swapped context—they actually showed a significant improvement ($F1(1,20) = 8.47$, $p < .05$; $F2(1,35) = 3.92$; $p < .05$).

The open-class tokens showed significantly higher accuracy relative to the neutral in both the original ($F1(1,20) = 6.14$, $p < .05$; $F2(1,35) = 7.41$, $p < .05$) and the swapped contexts ($F1(1,20) = 10.10$, $p < .05$; $F2(1,35) = 11.41$, $p < .05$). The closed-class tokens showed significant contextual facilitation (relative to the neutral contexts) in the original contexts ($F1(1,20) = 32.77$, $p < .05$; $F2(1,35) = 30.20$, $p < .05$), but not in the swapped contexts (F's < 1).

Open-class tokens are, essentially, recognized exceptionally well, yet still benefit from being placed in context; closed-class tokens, however, seem to fare poorly under non-‘‘normal’’ contexts, yet are recognized easily when in their original context. Further post-hoc analysis bears out these conclusions: the difference in accuracy between open- and closed-class targets was not significant in the original context ($F1(1,20) = 3.22$, $p = .088$; $F2(1,35) = 2.33$, $p = .136$), but was significant in both the neutral ($F1(1,20) = 78.40$, $p < .05$; $F2(1,35) = 21.59$, $p < .05$) and the swapped contexts ($F1(1,20) = 98.92$, $p < .05$; $F2(1,35) = 37.07$, $p < .05$).

The pattern of RTs was highly similar to that of error rates. Figure 2 demonstrates that what was difficult (low accuracy) corresponds roughly to what was repeated slowly. As Ta-

TABLE 1

ANALYSIS OF RT DATA FOR EXPERIMENT 1

	Subjects		Items	
	F	*df*	*F*	*df*
Context	6.20*	2,40	5.06*	2,35
Target	26.10*	1,20	22.30*	1,35
Context × Target	8.39*	2,40	3.69*	2,35
Contextual facilitation of RTs relative to neutral				
Original Open	5.58*	1,20	6.50*	1,35
Swapped Open	19.94*	1,20	17.86*	1,35
Original Closed	14.41*	1,20	4.53*	1,35
Swapped Closed	<1	1,20	<1	1,35
RT difference between open- and closed-class targets				
Original Context	1.70	1,20	2.52	1,35
Neutral Context	9.44*	1,20	4.26*	1,35
Swapped Context	28.45*	1,20	29.78*	1,35

* $p < .05$.

ble 1 shows, the main effects of and interaction between context and target were significant; the post-hoc comparisons that were significant in the accuracy data were also significant in the RT data.

Discussion

As expected, the closed-class targets seem to produce slower and less accurate responses. However, these differences disappear completely, for both accuracy and RT, when closed- and open-class items are compared in their original contexts. This indicates that closed-class words must, in some sense, rely to a greater extent on the presence of contextual cues; subjects are able to recognize closed-class tokens easily only when they appear in their normal context.

One criticism that might be leveled at the preceding experiment is that a ceiling of 1500 ms was used for the responses. Given that any responses occurring after the end of that window were considered omissions, there is a possibility that the accuracy data are not reflective of the actual difficulty of processing. We therefore tested a different group of 21 subjects on Experiment 1 with a longer RT window (3000 ms instead of 1500 ms, with all other conditions identical). Little effect of the length of the window was found. An ANOVA over the two groups showed no significant difference in RT ($F1 < 1$; $F2(1,35) = 2.26$, $p = .142$) or in accuracy ($F1(1,40) = 1.01$, $p = .321$; $F2(1,35) = 2.98$, $p = .093$). The most difficult cell (swapped-closed) showed the greatest effect of window length—responses grew relatively slower but more accurate—yet the overall ANOVA showed no significant Experiment × Context or Experiment × Target interaction. It therefore appears that a 1500-ms window for RTs forces subjects to perform very much in an on-line fashion without unnaturally affecting the speed or accuracy of their responses.

Thus the general finding from Experiment 1 is that listeners are more reliant on contextual information for the recognition of closed-class words. This does not explain, however, what kind of contextual information is contributing to the effect (given that ''context'' is a rather broad term). Although the length of the window used demands fairly rapid processing, it is possible that at least some of the contextual information following the target words is used in certain cases to assist in recognition of the already-heard target words. Subjects took, on the average, between 550 and 800 ms to recognize a word and respond to it. Subtracting the length of the target words (approximately 150 to 250 ms) leaves time for at least one or two of the following words to be heard. For the closed-class words it is possible that every bit of contextual information is helpful for processing. Experiment 2 addressed this question by forcing the subjects to base their responses on only the pretarget contextual information.

EXPERIMENT 2

Method

Subjects each received one of the three lists of 126 items described under General Methods. All trials were presented auditorily. The procedure administered to each subject was

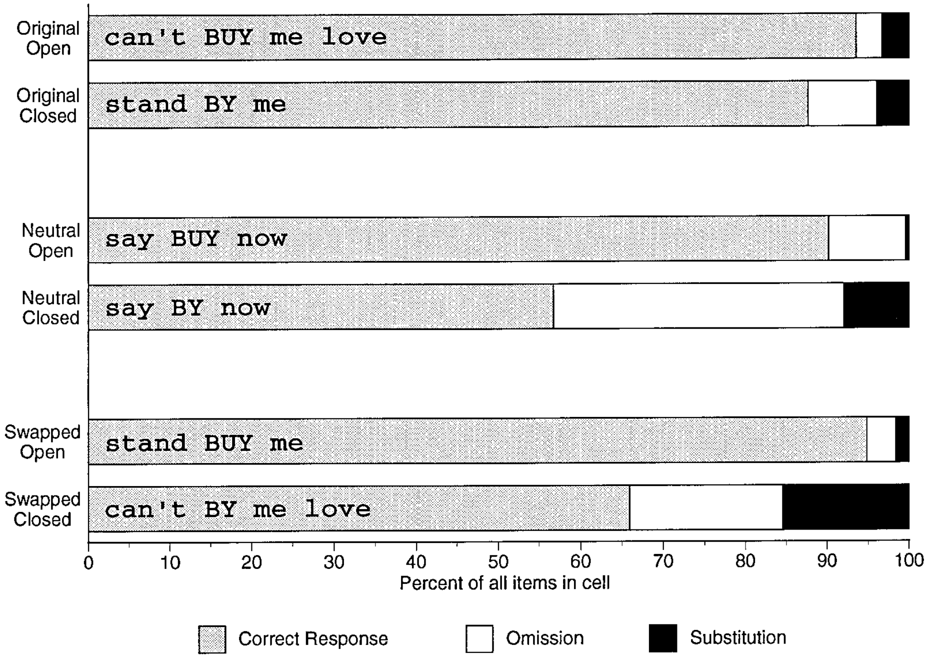

FIG. 3. Percentage of trials in Experiment 2 that were scored as correct, omitted, or substituted, as a function of target word class and sentence context. The example fragments are not representative of the actual stimuli.

identical to that under General Methods except that the portion of the sentence occurring *after* the target word was not presented until after the subject had finished responding or (if the subject failed to respond) 1500 ms had passed.

Results

Figure 3 shows the pattern of results for the accuracy data in Experiment 2. Significant main effects of context and target and a significant Context × Target interaction were present, as in Experiment 1. Closed-class items were significantly harder to identify; overall, subjects correctly identified 93% of the open-class targets and only 70% of the closed-class targets ($F1(1,20) = 236.13$, $p < .05$; $F2(1,35) = 21.71$, $p < .05$). The original contexts were easiest (91%), followed by the swapped contexts (80%), followed by the neutral contexts (73%) ($F1(2,40) = 16.47$, $p < .05$; $F2(2,35) = 14.05$, $p < .05$). A significant Context × Target interaction was also present ($F1(2,40) = 14.17$, $p < .05$; $F2(2,70) = 11.73$, $p < .05$).

The difference between open- and closed-class targets was not significant in the original context ($F1(1,20) = 2.94$, $p = .102$; $F2(1,35) = 2.03$, $p = .163$), but was significant in both the neutral ($F1(1,20) = 165.66$, $p < .05$; $F2(1,35) = 23.57$, $p < .05$) and the swapped contexts ($F1(1,20) = 44.04$, $p < .05$; $F2(1,35) = 18.14$, $p < .05$), as in Experiment 1.

The post-hoc measures of contextual facilitation were somewhat different than in Experiment 1, however. Specifically, responses to open-class tokens were not significantly more accurate in either the original ($F1(1,20) = 1.45$, $p = .243$; $F2 < 1$) or the swapped context ($F1(1,20) = 2.75$, $p = .113$; $F2(1,35) = 1.71$, $p = .200$), relative to the neutral context. The closed-class tokens showed significant contextual facilitation in the original contexts ($F1(1,20) = 36.61$, $p < .05$; $F2(1,35) = 29.24$, $p < .05$), but not in the swapped contexts ($F1(1,20) = 2.55$, $p = .126$; $F2(1,35) = 1.48$, $p = .231$).

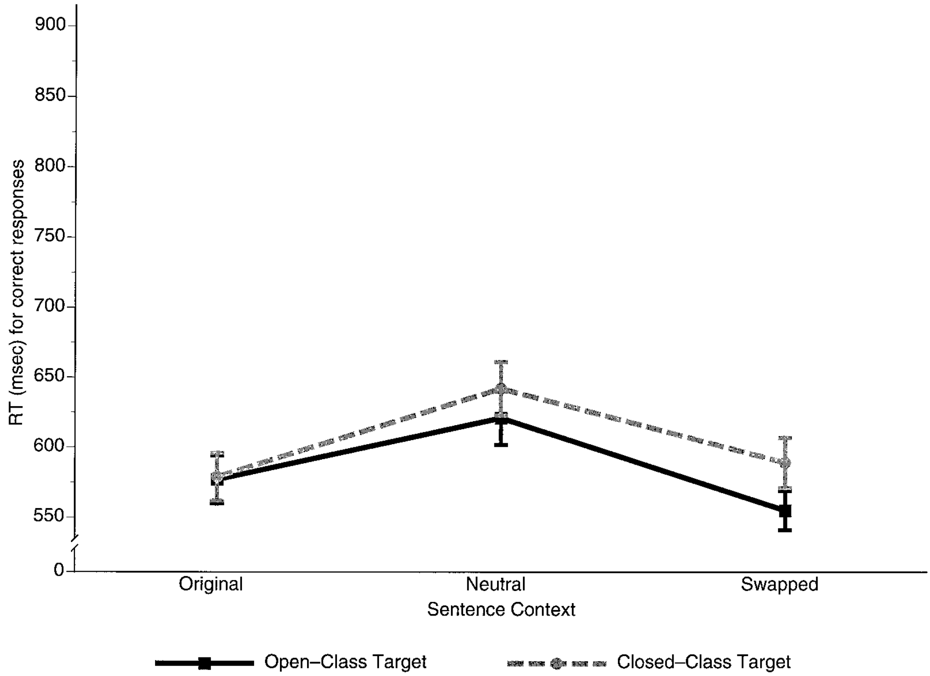

FIG. 4. Mean RT across subjects to correct responses in Experiment 2, as a function of target word class and sentence context. The error bars represent the standard error across subjects.

To compare directly Experiments 1 and 2, ANOVAs were performed with experiment, context, and target as factors. No main effect of experiment on accuracy was found, nor any interaction of experiment with other factors.

The pattern of RTs was somewhat similar to that of the accuracy data, as shown in Fig. 4. The Context × Target interaction, however, did not reach significance, and the main effect of target was marginally significant. The post-hoc comparisons differed in several cases from those for the accuracy data, as detailed for the interested reader in Table 2.

Discussion

Experiment 2 addressed the question of the degree to which the context following the target word aided in recognition of that word; the effect of the post-target contextual information should be visible as the difference in performance between Experiment 1 (both pre- and post-target context) and Experiment 2 (pretarget context only). Very little difference was observed in the accuracy of responses. Experiment 2 did, however, yield a relative flattening of RT effects. This includes smaller contextual facilitation effects across the two word classes, and a smaller RT difference between open- and closed-class words (reliable over items and subjects in the swapped condition, but reliable only over items in the neutral condition).

The differences between Experiments 1 and 2 were relatively small, though, indicating that the context following the target word has a minor effect on recognition (otherwise one would have expected greater differences in accuracy between Experiments 1 and 2). At the very least, this means that the context effects and class differences observed in Experiment 1 were not due solely to post-target context. In fact, performance actually became faster when subjects were deprived of post-target context, perhaps because the subjects in Experiment 2 enjoyed a lessened processing load and/or less acoustic masking of the stimulus

TABLE 2

ANALYSIS OF RT DATA FOR EXPERIMENT 2

	Subjects		Items	
	F	*df*	*F*	*df*
Context	26.45*	2,40	15.19*	2,35
Target	3.99	1,20	9.09*	1,35
Context × Target	1.42	2,40	1.37	2,35
Contextual facilitation of RTs relative to neutral				
Original Open	20.80*	1,20	2.92	1,35
Swapped Open	54.42*	1,20	34.42*	1,35
Original Closed	20.86*	1,20	11.52*	1,35
Swapped Closed	7.76*	1,20	8.02*	1,35
RT difference between open- and closed-class targets				
Original Context	<1	1,20	<1	1,35
Neutral Context	1.98	1,20	5.20*	1,35
Swapped Context	4.93*	1,20	7.44*	1,35

* $p < .05$.

(since they heard silence until they responded), compared with subjects in Experiment 1. This could also explain why the RTs to the swapped-closed items fell in Experiment 2. Perception of the closed-class targets was no longer masked by any ongoing sentence; the subjects did not have to try to process additional incoming information and were thus free to concentrate their resources on processing the closed-class target. Closed-class words, which are in general less salient, may be easier to pick out from the acoustic stream when they are not masked by any following words. Open-class words, however, are far more salient and may not show the same benefit.

Thus our basic conclusions from Experiment 1 are confirmed; closed-class words seem to be more dependent on contextual information. This can be further analyzed, however; ‘‘contextual information’’ can be broken down into different forms of contextual information. Of particular relevance to the open/closed distinction may be prosodic information. Prosodic information can provide listeners with some cues to upcoming words. Cutler and Foss (1977), for instance, showed that preceding prosodic information is predictive of the high stress given to the focus of a sentence (which is almost always open-class). There is reason to suspect that the prosodic information that is present not only within but before the target word can influence processing and, more importantly, to suspect that such prosodic information may have very different effects on open-class versus closed-class words. The goal of Experiment 3 was to examine the role of prosodic information in the processing of our stimuli. To accomplish this, the information preceding (and following) the target word was presented without prosodic information (visually); thus any differences in processing between Experiments 1 and 3 should reflect the contribution of prosody.

EXPERIMENT 3

Method

Subjects each received one of the three lists of 126 items described under General Methods. The procedure in this experiment deviated from that described under General Methods in that presentation of all words except the target was on a computer monitor. Each word was displayed on a computer monitor approximately 24 inches in front of the subject for 300 ms, with 150 ms between words. The target word was played over headphones 150 ms after the visual offset of the preceding word; the post-target context was presented visually, beginning 150 ms after the end of the (auditory) target word.

Results

Figure 5 presents the accuracy data for Experiment 3. Responses for original contexts (83%) were more accurate than those for swapped (79%) and neutral contexts (72%) ($F1(2,40) = 4.32$, $p < .05$; $F2(2,70) = 8.82$, $p < .05$). There was significantly higher accuracy for open-class targets (90%) than closed-class (66%) targets ($F1(1,20) = 161.44$, $p < .05$; $F2(1,35) = 22.53$, $p < .05$). The Context × Target interaction was also significant

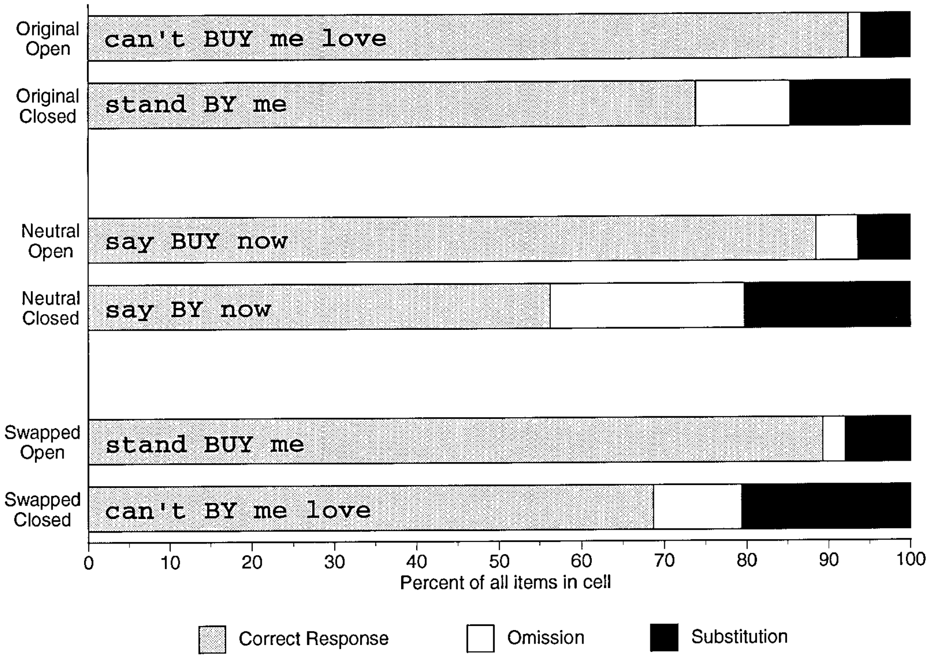

FIG. 5. Percentage of trials in Experiment 3 that were scored as correct, omitted, or substituted, as a function of target word class and sentence context. The example fragments are not representative of the actual stimuli.

($F1(2,40) = 3.28$, $p < .05$; $F2(2,70) = 3.78$, $p < .05$).

A large numerical drop in accuracy for the closed-class targets in their original context (relative to the performance in Experiment 1) can be seen in comparing Figs. 1 and 5. This is emphasized by noting that the difference in accuracy between open- and closed-class targets was significant not only in the neutral ($F1(1,20) = 73.89$ $p < .05$; $F2(1,35) = 27.19$, $p < .05$) and swapped contexts ($F1(1,20) = 24.95$, $p < .05$; $F2(1,35) = 10.98$, $p < .05$), but in the original ($F1(1,20) = 29.30$, $p < .05$; $F2(1,35) = 12.98$, $p < .05$) as well.

Responses to open-class tokens were not significantly more accurate in either the original ($F1(1,20) = 1.51$, $p = .234$; $F2(1,35) = 1.00$, $p = .324$) or the swapped context (F's < 1), relative to the neutral context. The results did show, however, significant contextual facilitation of the closed-class tokens in the original context ($F1(1,20) = 9.05$, $p < .05$; $F2(1,35) = 26.76$, $p < .05$), however, indicating that the semantic/syntactic context that remained under the visual presentation was still of benefit to the subjects. The closed-class targets in the swapped context showed only marginal contextual facilitation ($F1(1,20) = 4.10$, $p = .056$; $F2(1,35) = 4.94$, $p < .05$).

To compare directly Experiments 1 and 3, ANOVAs were performed with experiment, context, and target as factors. Subjects were, in general, less accurate in Experiment 3; the main effect of experiment was significant ($F1(1,40) = 5.52$, $p < .05$; $F2(1,70) = 7.50$, $p < .05$). The greatest drop in accuracy was for the responses to the closed-class targets in their original context (which fell from approximately 86% to 74%). The Experiment $\times$ Context interaction was significant over items ($F1(2,80) = 2.73$, $p = .071$; $F2(2,70) = 4.81$, $p < .05$), the Experiment $\times$ Target interaction was not significant (F's < 1), and the Experiment $\times$ Context $\times$ Target interaction was significant ($F1(2,80) = 5.24$, $p < .05$; $F2(2,70)$

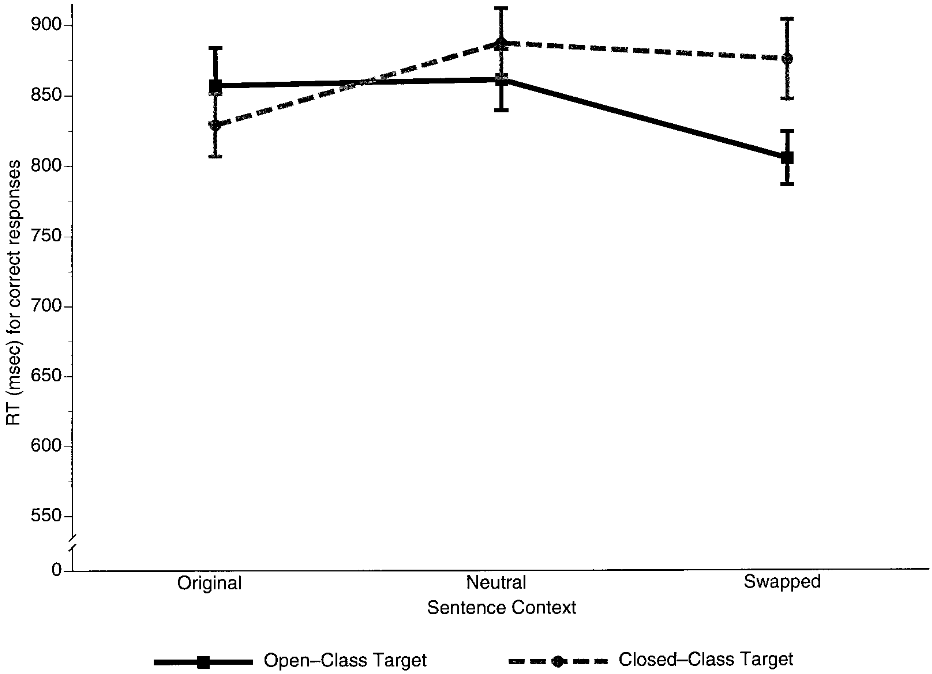

FIG. 6. Mean RT across subjects to correct responses in Experiment 3, as a function of target word class and sentence context. The error bars represent the standard error across subjects.

= 3.90, $p < .05$). To analyze this effect further, we considered only the targets in their original contexts; an ANOVA showed a significant Experiment × Target interaction ($F1(1,40) = 11.33$, $p < .05$; $F2(1,35) = 5.62$, $p < .05$). The same interaction was not significant in the neutral or the swapped contexts. This significant interaction in the original context reflects the fact that accuracy to open- and closed-class targets was nearly equal in Experiment 1, while closed-class performance was far worse than open-class performance in this experiment. In other words, prosodic information seems to be especially important for closed-class words in their original context.

The pattern of RTs was somewhat similar to that of the accuracy data, as shown in Fig. 6. The main effects of context and target and the Context × Target interaction were significant. The post-hoc comparisons differed in several cases from those for the accuracy data, as seen in Table 3.

Discussion

Visual presentation of the contextual information had a clear effect on processing, relative to Experiment 1. RTs were slower and accuracy was also affected, but more so for the closed-class words. In Experiment 1, when subjects were processing sentences with normal prosodic information, responses to the closed-class targets were highly accurate when those targets were embedded in their original contexts. In the absence of prosodic information, recognition of closed-class targets dropped significantly, but only in the cell in which performance had previously been good (the original context).

This result leads us to suggest that closed-class items are ‘‘prosody dependent.’’ Although closed-class words are less salient overall under normal listening conditions (partly because they generally contain only weak syllables), listeners take advantage of the prosodic envelope in which those words are embedded as an aid to recognition. If this

TABLE 3

ANALYSIS OF RT DATA FOR EXPERIMENT 3

	Subjects		Items	
	F	*df*	*F*	*df*
Context	3.94*	2,40	4.96*	2,35
Target	4.52*	1,20	10.10*	1,35
Context × Target	9.15*	2,40	3.92*	2,35
Contextual facilitation of RTs relative to neutral				
Original Open	<1	1,20	<1	1,35
Swapped Open	29.13*	1,20	14.02*	1,35
Original Closed	15.70*	1,20	6.79*	1,35
Swapped Closed	<1	1,20	1.49	1,35
RT difference between open- and closed-class targets				
Original Context	5.71*	1,20	<1	1,35
Neutral Context	3.31	1,20	7.02*	1,35
Swapped Context	9.23*	1,20	10.22*	1,35

* $p < .05$.

interpretation is correct, it may help to explain a puzzling finding in the literature on oral and written language abilities in the congenitally deaf. It has been known for some time that grammatical function words pose a serious problem for deaf individuals who acquire an oral language (reflected in difficulties in using the proper function word at the proper time). This is true even for those individuals who achieve very high levels of lexical and syntactic proficiency (Volterra & Bates, 1989). At first, one might assume that this selective deficit reflects the fact that closed-class words are low in acoustic salience; but this cannot be the explanation, because all words are low in acoustic salience for the profoundly deaf. Indeed, this is the one group for whom relative differences in salience should not matter at all. If, however, recognition of closed-class words is prosody dependent in the hearing and the deaf do not have access to prosody during language learning and/or language use, then the deaf might find it difficult to achieve native levels of proficiency in the processing of grammatical function words.

Although we have concluded that prosody does play a role in recognition of closed-class words, the semantic/syntactic context that remains under visual presentation in Experiment 3 is also important. That is, closed-class words are still more context dependent than open-class words in this experiment. It appears that subjects are able to integrate prosodic, semantic, and syntactic cues to overcome the low acoustic salience of closed-class words.

What is the basis of the acoustic difference between open- and closed-class words? As we have noted, closed-class words generally have only weak syllables, while open-class words almost always have at least one strong syllable (Cutler, 1993). Strong syllables are those that have unreduced vowels and weak syllables are those that have reduced vowels, such as schwa. Vowel reduction is not the only way that syllable strength is marked, however; syllabic /r̩/, /l̩/, and /n̩/ can also take the place of the vowel as the nucleus (Dauer, 1983), for example *or* → /r̩/ in the stimuli used in these experiments. Furthermore, other perceptible differences aside from vowel quality contribute to making open-class words more salient or stressed than closed-class words—some closed-class words have unreduced vowels, yet are still less salient than their open-class mates.

All but one of the target word-pairs used in these experiments were one syllable in length (the exception being *weather*/*whether*), and so for the purposes of this paper the open- and closed-class targets will be referred to as stressed and unstressed words (because they are only one syllable long, any stress must by definition be primary stress). The stress differences between open- and closed- class words are often described acoustically as differences in length, amplitude, pitch, and vowel structure. Stress is difficult to measure, however, and it is unclear what the most important acoustic variables are; both duration (Fry, 1955) and pitch (Bolinger, 1958; Fry, 1958) have been claimed as the main acoustic correlates of stress. Duration and amplitude are also claimed to serve as cues to grammatical class (Sereno & Jongman, 1995). The duration of the target words used in these experiments did show a significant difference in length: open-class words were far longer on average (265 ms) than were closed-class words (167 ms) ($F(1,35) = 60.75$, $p < .05$). However, no significant difference between the open- and closed-class targets was found for RMS amplitude or pitch (pitch is, however, difficult to measure reliably, especially for very short segments like the closed-class targets).

Because the processing differences observed between the open- and closed-class words must be dependent in some way on acoustic differences (otherwise the system could not distinguish between the word classes), one could reasonably imagine that such gross differences in length might play a part in the processing differences observed between open- and closed-class words. Furthermore, length is not only a very salient property, but also one that is relatively easy to alter, thus allowing us to separate out its effect from other effects (ideally, however, one would like to test the role of the other prosodic properties on processing). In order to try to determine what role this particular acoustic difference played in the word class effects obtained in Experiments 1–3, in Experiment 4 the length difference between open- and closed-class targets was removed.

EXPERIMENT 4

Method

Subjects each received one of the three lists of 126 items described under General Methods, with the target words digitally manipulated so that the lengths of both members of each homophone pair were equal. The target words were digitally lengthened or shortened toward the mean within each homophone pair; for example, *dew* was shortened and *do* was lengthened so that the two new targets had the same length. The SoundEdit 16 software package for the Macintosh computer was used to perform this manipulation; the length-altering algorithm looks for repeated sections of the stimulus to remove or duplicate in order to shorten or lengthen the sound. (Note that although differences in vowel length are probably the major component of the word-length differences, the manipulation used here was applied to the entire target word, due to the difficulty of locating only the vocalic segments.) All trials were presented auditorily, using the combined headphones and microphone. The procedure administered to each subject was identical to that described under General Methods.

Results

Figure 7 shows the pattern of responses in Experiment 4. The pattern of results was highly similar to that found in Experiment 1. Responses for original contexts (91%) were more accurate than those for swapped (83%) and neutral contexts (70%) ($F1(2,40) = 15.58$, $p < .05$; $F2(2,70) = 19.69$, $p < .05$). Responses to open-class targets were significantly more accurate (91%) than those to closed-class (71%) targets ($F1(1,20) = 152.44$, $p < .05$; $F2(1,35) = 16.34$, $p < .05$). The Context $\times$ Target interaction was significant ($F1(2,40) = 10.29$, $p < .05$; $F2(2,70) = 5.69$, $p < .05$).

The open-class items showed significant contextual facilitation in both the original ($F1(1,20) = 6.49$, $p < .05$; $F2(1,35) = 7.57$, $p < .05$) and the swapped ($F1(1,20) = 5.57$, $p < .05$; $F2(1,35) = 9.66$, $p < .05$) contexts. The closed-class items also showed contextual facilitation in both the original ($F1(1,20) = 32.12$, $p < .05$; $F2(1,35) = 23.39$, $p < .05$) and the swapped contexts ($F1(1,20) = 8.08$, $p < .05$; $F2(1,35) = 10.03$, $p < .05$).

The difference in accuracy between open- and closed-class items was significant in the neutral ($F1(1,20) = 93.75$, $p < .05$; $F2(1,35) = 14.14$, $p < .05$) and the swapped conditions ($F1(1,20) = 33.53$, $p < .05$; $F2(1,35) = 14.71$, $p < .05$). This difference was significant only over subjects in the original context ($F1(1,20) = 5.56$, $p < .05$; $F2(1,35) = 2.33$, $p = .136$).

The similarity in performance between Experiments 1 and 4 is emphasized by the fact that there was no main effect of experiment nor interaction of experiment with context and/or target in the ANOVA comparing the two experiments.

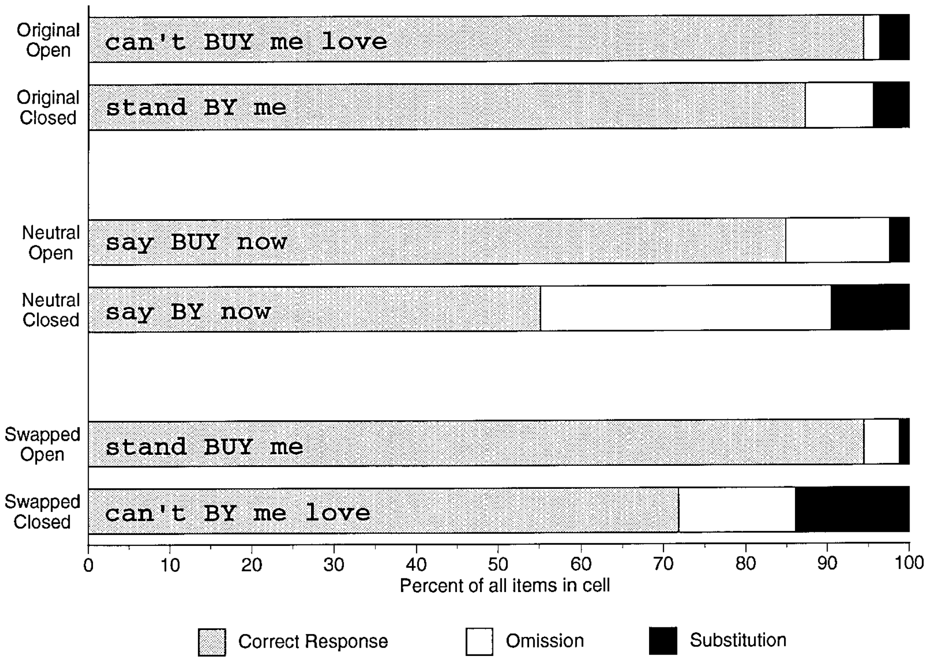

FIG. 7. Percentage of trials in Experiment 4 that were scored as correct, omitted, or substituted, as a function of target word class and sentence context. The example fragments are not representative of the actual stimuli.

The pattern of RTs (seen in Fig. 8) was somewhat similar to that of error rates. As Table 4 shows, the main effects of and interaction between context and target were significant (although the main effect of context was not significant over items). The post-hoc comparisons of the RT data differed from those for the accuracy data, however.

Discussion

The failure to find a significant difference with Experiment 1 in either accuracy or RTs stemming from the use of the length-altered stimuli suggests that length is not the major acoustic cue to the stress differences that distinguish open- and closed-class words. It is, however, difficult to define stress acoustically; ambiguity and variation seem to be the hallmarks of experimental investigation of stress. The failure to find an effect of duration in the present experiments is nonetheless surprising, given the apparently large difference between the target words. There were subtle traces of increased accuracy for the lengthened closed-class words (as would be hypothesized if length were an important variable), but equating the lengths of the open- and closed-class words by no means altered the general finding that closed-class words suffer more than open-class words when in the swapped and neutral contexts. Presumably other factors, such as vowel quality and fundamental frequency, are responsible for the perceived and relevant difference between the open and closed classes.

Vowel differences are, in fact, claimed to distinguish between stressed and unstressed syllables. Fear et al. (1995) found that listeners appear to distinguish between strong and weak syllables based on whether or not the syllable has a reduced vowel, while unstressed but unreduced vowels are grouped with stressed vowels. (Note that the closed-class tokens used in this experiment do not all have reduced vowels. Although it is difficult to phonetically transcribe words of such short duration, it appears that only a third of the closed-

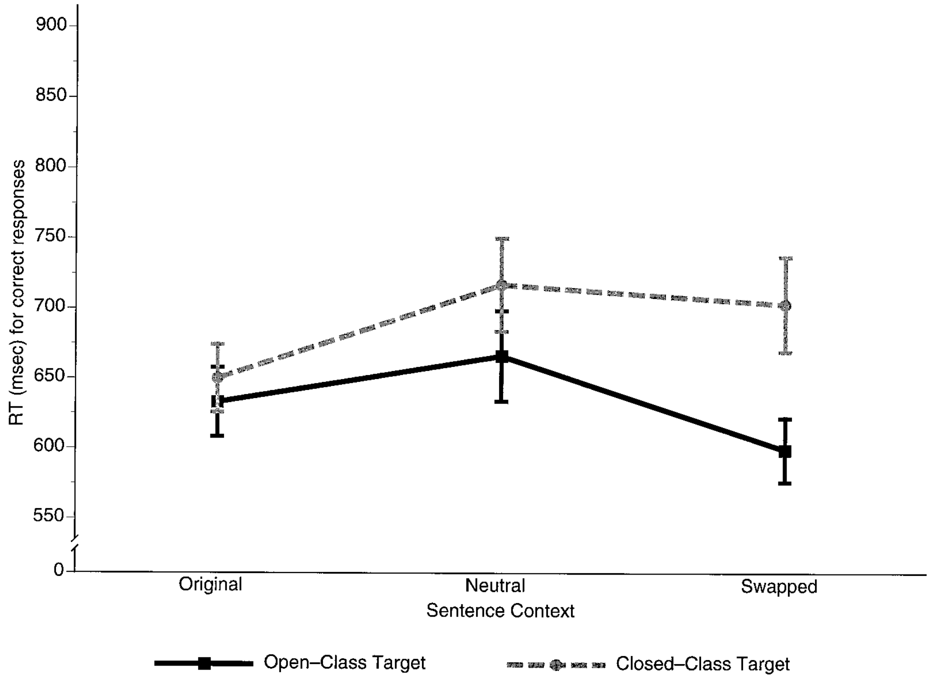

FIG. 8. Mean RT across subjects to correct responses in Experiment 4, as a function of target word class and sentence context. The error bars represent the standard error across subjects.

class tokens had reduced vowels. In other cases the vowel was distinct from the open-class pronunciation, but not reduced. In yet other cases, the word contained either a tense vowel or a diphthong, neither of which reduce. Several of the closed-class tokens had syllabified consonants in place of any vowel at all.)

At the very least, however, these results mean that prelexical sorting of words by class cannot be based on a single, simple cue like length. If there is a filter or sorting mechanism of some kind that operates prior to word recognition, then that mechanism must be sensitive to many smaller variations in phonetic structure, deformations that may vary on a word-by-word basis.

GENERAL DISCUSSION

Closed-class items can be said to have a dual status in processing. Because they are highly frequent and often contextually constrained, they should be relatively easy to process; they are, however, generally acoustically degraded, making them difficult to process. It has long been known that closed-class words suffer when taken out of context (Pollack & Pickett, 1964) and it may not be surprising, then, that these experiments demonstrated such a high degree of contextual dependence for the closed class. One must keep in mind, however, that open-class words also benefit from their context, although to nowhere near the degree that closed-class words do. It also makes sense that both word classes are processed easily when in their normal contexts; as Cutler (1993) pointed out, however different the word classes may be, it would be odd to discover that closed-class words cause real processing difficulties, since they are so frequent. (Of course, this holds true only in the normal case; closed-class items are notoriously sensitive to pathology.)

Why is it that closed-class words have such a reduced acoustic status? As these experiments have shown, stressed words are uniformly easy to process; why, then, do we not

TABLE 4

ANALYSIS OF RT DATA FOR EXPERIMENT 4

	Subjects		Items	
	F	*df*	*F*	*df*
Context	5.37*	2,40	<1	2,35
Target	51.89*	1,20	20.64*	1,35
Context × Target	5.06*	2,40	3.71*	2,35
Contextual facilitation of RTs relative to neutral				
Original Open	3.73	1,20	<1	1,35
Swapped Open	14.61*	1,20	3.81	1,35
Original Closed	6.29*	1,20	<1	1,35
Swapped Closed	<1	1,20	1.86	1,35
RT difference between open- and closed-class targets				
Original Context	1.25	1,20	<1	1,35
Neutral Context	12.81*	1,20	3.36	1,35
Swapped Context	21.35*	1,20	17.53*	1,35

* $p < .05$.

stress all words? The obvious and perhaps circular answer is that closed-class words *can* depend on contextual information so highly, thus allowing the speaker to save time and energy in production. Grosjean and Gee (1987) made the claim that reduction in response to contextual constraint will take place even with open-class words (when the context is sufficiently constraining). Speakers do, however, sometimes wish to emphasize closed-class words for pragmatic reasons, and thus closed-class words are sometimes uttered in their canonical form.

This fact alone ought to doom, on theoretical grounds, the claim that the processing of open- and closed-class words is divided based solely on their bottom-up (acoustic) properties—since closed-class words sometimes sound like open-class words, such a simple heuristic will lead to unacceptably many failures. In these experiments, stressing closed-class words actually seemed to improve recognition (accuracy in the swapped-open cell was numerically higher than in the original-closed cell in all four experiments, as visible in Figs. 1, 3, 5, and 7), which contradicts the predictions of even those models in which closed-class words are represented in both lexica or both pathways. Furthermore, even word duration, a seemingly strong bottom-up predictor of word-class, was found to have very little influence on processing. Thus of the three models outlined in the introduction, only the second two (the ''limited modularity'' and the interactive models) remain plausible. Unfortunately, distinguishing empirically between these alternatives may be impossible—there is no well-defined explanation of what lexical access is or when it occurs, and both alternatives are poorly defined and subject to ad hoc revisions that allow them to fit any set of results. It may even be theoretically impossible to distinguish between the two models (Altmann & Steedman, 1988).

Regardless, one can argue that the simplest explanation of the nonadditive effects of contextual and acoustic information found in these experiments is symptomatic of the simultaneous use of multiple cues in the access of a particular lexical entry. The mapping of these cues onto a lexical target appears to happen relatively quickly; subjects made little use of the contextual information following the target words, indicating that they were processing very much in an ''on-line'' fashion (rather than guessing post-hoc at the identity of each word.) When one accounts for the time taken to perceive the target word and to activate the motor code for the response, the RTs in these experiments lie at least within the ballpark of lexical access (keeping in mind how poor an understanding we have of when this phenomenon takes place).

Furthermore, we have shown that ''context'' refers, in this case, not simply to the semantic or grammatical constraints imposed by previous words, but to the prosodic structure of the surrounding information. In the interactive view, such prosodic information is merely one more cue that predicts (however weakly) the lexical identity of the target word in just the same manner as any other contextual or acoustic cue. The failure to find a major effect of target word length on processing perhaps indicates that many different aspects of the acoustic information are integrated to achieve recognition, with each having a variable strength (e.g., something other than length, such as vowel differences, may be more relevant to the difference between open- and closed-class words). Again, a so-called weakly modular model (in which prosodic information is processed in parallel with but separately from other information) can fit these data; our argument is that parsimony would argue for a single processor that handles all information simultaneously.

How does the interactive model predict the poor performance with the swapped-closed items? This could be due to a lack of ecological validity for those items: Open-class items are always stressed, while closed-class items are often heard both stressed and unstressed. Therefore the processor will be able to map the input onto the proper entry in every case except for the unstressed open-class item. This reply, of course, borders on question-begging, as it does not provide a satisfactory explanation of why those acoustic differences exist. As discussed above, however, one might argue that the desire of the speaker to articulate quickly, combined with his or her knowledge (as a listener) that closed-class words are highly predictable, might lead to the solution of reduced closed-class words. A second but related response draws on Cutler's (1993; see also Cutler & Carter, 1987) claim that a segmentation strategy based on strong syllables can facilitate lexical access. Having certain words stand out acoustically makes sense, as it can allow the system to find boundaries more easily. Furthermore, that those boundaries are often defined in a way that distinguishes open- and closed-class words is reasonable, but not because this distinction is used to then process those words in separate processors. Rather, the acoustic distinction can serve as just one more cue to the identity of the word, just as featural and/or phonological information is used to distinguish among many words in the lexicon. Similarly, the distinction between other word classes, such as nouns and verbs, may be indicated acoustically (Sereno & Jongman, 1990, 1995).

This cued shadowing task does not tell us enough about the time course or the nature of processing to determine really what it is about the swapped-closed items that presents difficulty. Nevertheless, the interaction of acoustic and contextual information is not explicable solely in terms of the acoustic weakness of the closed-class targets—in the proper context, both the salient and the less salient tokens can be easily and quickly processed. The degree and type of contextual information also seem not to fully explain the results—both contexts show that they can facilitate processing of the target words. Ideally, one would like to find a task that could better determine the nature of the interaction between these cues (especially when and how they are integrated.) Without innovative techniques and a better specification of the various theories of the processing of open- and closed-class words, however, it is impossible to empirically choose among them.

Meanwhile, our results do have some interesting implications for theories of language processing under abnormal conditions (Bates, Wulfeck, & MacWhinney, 1991; Blackwell & Bates, 1995; Frazier & Friederici, 1991; Kilborn, 1991; Miyake, Carpenter, & Just, 1994). It has been argued that closed-class words constitute a ''weak link in the processing chain,'' although perhaps only under abnormal conditions. That is, these words appear to be especially vulnerable (at least in receptive language processing) in aphasia, and they are also selectively and disproportionately impaired when normal subjects are forced to process sentences under a partial noise mask (Kilborn, 1991), dual-task conditions (Blackwell & Bates, 1995), and/or speeded processing (Bates, Devescovi, Dronkers, Pizzamiglio, Wulfeck, Hernandez, Juarez, & Marangolo, 1994). If it is the case (as we have argued here) that efficient and accurate processing of closed-class words is highly dependent on context, then any condition that disrupts the timing and integration of acoustic and contextual information will have selectively greater effects on closed-class words.

APPENDIX

TARGET WORDS USED IN THE EXPERIMENTAL ITEMS

Closed-class targets		Open-class targets	
Target	Pronunciation	Target	Pronunciation
be	/bi/	bee	/bi/
but	/bʌt/	butt	/bʌt/
by	/bʌ/	buy	/baɪ/
by	/baɪ/	bye	/baɪ/
can	/kɪn/	can	/kæn/
do	/du/	dew	/du/
do	/du/	due	/du/
down	/daɰn/	down	/daɰn/
for	/fr̩/	four	/fɔr/
have	/hæf/	halve	/hæv/
him	/ɪm/	hymn	/hɪm/
I	/e/	eye	/aɪ/
in	/ɪn/	inn	/ɪn/
may	/meɪ/	May	/meɪ/
might	/maɪt/	might	/maɪt/
might	/maɪt/	mite	/maɪt/
mine	/maɪn/	mine	/maɪn/
must	/məs/	must	/mʌst/
need	/nju/	knead	/nid/
no	/no/	know	/no/
none	/nʌn/	nun	/nən/
not	/nat/	knot	/nɒt/
or	/r̩/	oar	/ɔr/
our	/ər/	hour	/aɰr/
so	/si/	sew	/so/
some	/səm/	sum	/səm/
to	/t/	two	/tu/
too	/tu/	two	/tu/
we	/wi/	wee	/wi/
where	/wər/	wear	/wɛr/
whether	/wɛðr̩/	weather	/wɛðr̩/
while	/wʌ/	wile	/waɪl/
will	/ʌl/	will	/wɪl/
which	/wɪč/	witch	/wɪč/
would	/wud/	wood	/wɰd/
you	/ju/	ewe	/ju/

REFERENCES

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition,* **30,** 191–238.

Bates, E., Devescovi, A., Dronkers, N., Pizzamiglio, L., Wulfeck, B., Hernandez, A., Juarez, L., & Marangolo, P. (1994). Grammatical deficits in patients without agrammatism: Sentence Interpretation under stress in English and Italian. *Brain and Language,* **47,** 400–402.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). New York: Cambridge Univ. Press.

Bates, E., & Wulfeck, B. (1989). Crosslinguistic studies of aphasia. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 328–371). New York: Cambridge Univ. Press.

Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Crosslinguistic studies in aphasia: An overview. Special Issue: Crosslinguistic studies of aphasia. *Brain and Language,* **41,** 123–148.

Blackwell, A., & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience,* **7,** 228–257.

Bolinger, D. L. (1958). A theory of pitch accent. *Word,* **14,** 109–149.

Bradley, D. C. (1978). *Computational distinctions of vocabulary type.* Unpublished doctoral dissertation, MIT.

Bradley, D. C., & Garrett, M. F. (1983). Hemisphere differences in the recognition of closed and open class words. *Neuropsychologia,* **21,** 155–159.

Chiarello, C., & Nuding, S. (1987). Visual field effects for processing content and function words. *Neuropsychologia,* **25,** 539–548.

Church, K. W. (1987). Phonological parsing and lexical retrieval. Special Issue: Spoken word recognition. *Cognition,* **25,** 53–69.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments and Computers,* **25,** 257–271.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105–121). Cambridge, MA: MIT Press.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. Special Issue: Sentence processing: III. *Journal of Psycholinguistic Research,* **22,** 109–131.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language,* **2,** 133–142.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech,* **20,** 1–10.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics,* **11,** 51–62.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America,* **97,** 1893–1904.

Frazier, L., & Friederici, A. (1991). On deriving the properties of agrammatic comprehension. *Brain and Language,* **40,** 51–66.

Friederici, A. D. (1982). Syntactic and semantic processes in aphasic deficits: The availability of prepositions. *Brain and Language,* **15,** 249–258.

Friederici, A. D. (1985). Levels of processing and vocabulary types: Evidence from on-line comprehension in normals and agrammatics. *Cognition,* **19,** 133–166.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America,* **27,** 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech,* **1,** 126–152.

Garrett, M. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 9, pp. 133–177). New York: Academic Press.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. Special Issue: Spoken work recognition. *Cognition,* **25,** 135–155.

Kilborn, K. (1991). Selective impairment of grammatical morphology due to induced stress in normal listeners: Implications for aphasia. *Brain and Language,* **30,** 580–602.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown Univ. Press.

Liu, H., Bates, E., Powell, T., & Wulfeck, B. (in press). Single-word shadowing and the study of lexical access. *Applied Psycholinguistics.*

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676–703.

Matthei, E. H., & Kean, M. L. (1989). Postaccess processes in the open vs. closed class distinction. *Brain and Language,* **36,** 163–180.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology,* **11,** 671–717.

Petocz, A., & Oliphant, G. (1988). Closed-class words as first syllables do interfere with lexical decisions for nonwords: Implications for theories of agrammatism. *Brain and Language,* **34,** 127–146.

Pollack, I., & Pickett, J. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural content. *Journal of Verbal Learning and Memory,* **3,** 79–84.

Sereno, J. A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research,* **19,** 387–404.

Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech,* **38,** 57–76.

Shapiro, L. P., & Jensen, L. R. (1986). Processing open and closed class-headed nonwords: Left hemisphere support for separate vocabularies. *Brain and Language,* **28,** 318–327.

Shillcock, R. C., & Bard, E. G. (1993). Modularity and the processing of closed class words. In G. T. M. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 163–183). Hove, UK: Erlbaum.

Swinney, D. (1982). The structure and time-course of information interaction during speech comprehension: lexical segmentation, access, and interpretation. In J. Mehler, E. C. T. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representation: experimental and theoretical studies of cognitive processes and capacities* (pp. 151–165). Hillsdale, NJ: L. Erlbaum.

Swinney, D. A., Zurif, E. B., & Cutler, A. (1980). Effects of sentential stress and word class upon comprehension in Broca's aphasics. *Brain and Language,* **10,** 132–144.

Tanenhaus, M. K., & Lucas, M. M. (1987). Context effects in lexical processing. Special Issue: Spoken word recognition. *Cognition,* **25,** 213–234.

Volterra, V., & Bates, E. (1989). Selective impairment of Italian grammatical morphology in the congenitally deaf: A case study. *Cognitive Neuropsychology,* **6,** 273–308.

(Received December 26, 1995)
(Revision received February 12, 1997)