BRAIN AND LANGUAGE 15, 143–160 (1982)

Lexical Decision for Open- and Closed-Class Words: Failure to Replicate Differential Frequency Sensitivity

BARRY GORDON AND ALFONSO CARAMAZZA

The Johns Hopkins University School of Medicine and The Baltimore City Hospitals

Bradley and her colleagues (D. C. Bradley, *Computational distinctions of vocabulary type*, unpublished Ph.D. thesis, Massachusetts Institute of Technology, 1978; D. C. Bradley, M. E. Garrett, & E. B. Zurif, in D. Caplan (Ed.), *Biological studies of mental processes*, Cambridge, MA: MIT Press, 1980) have reported a marked difference in frequency sensitivity between open- and closed-class words on a lexical decision task. This effect was obtained with normal subjects, but not with Broca's aphasics. Their results have already influenced experimental and theoretical investigations of syntactic processing. However, in three lexical decision experiments with normal subjects, modeled on those of Bradley et al., we failed to find such a theoretically interesting difference between the two classes. Instead, both classes showed similar reaction time frequency sensitivity for word frequencies less than approximately 316/million (H. Kucera & W. N. Francis, *Computational analysis of present-day English*, Providence, RI: Brown Univ. Press, 1967, count); above 399/million, the closed class had an almost-flat function of reaction time versus the logarithm of the frequency, while the open class may have had too few members for meaningful assessment. Because reaction time may be a nonlinear function of log frequency, and because there is relatively little overlap between the frequency ranges of the two classes, comparisons of the members of the two classes which might straddle the function's inflection point must be made with extreme caution.

Word frequency has been shown to have a pervasive influence on a multitude of psychological processes. For example, in lexical decision tasks, word frequency has a reliable negative correlation with latencies

This study was supported in part by Grants NS 16155 and NS 14099 from the National Institutes of Health (NINCDS) to The Johns Hopkins University. The authors wish to thank Dr. Steven G. Lapointe for help in selecting stimuli, Kevin Gallagher and John P. Lewis for programming, Clara Marin and Andrew M. Mead for testing subjects, and Dr. Rita Berndt, Dr. Lapointe, and Renee Gordon for a critical review of the manuscript. We also thank Drs. Merrill Garrett, Edgar Zurif, and an anonymous reviewer for their helpful comments on portions of this work. Address reprint requests to Barry Gordon, M.D., Ph.D., Department of Neurology, Baltimore City Hospitals, 4940 Eastern Avenue, Baltimore, MD 21224.

0093-934X/82/010143-18$02.00/0

for correct responses to word stimuli (e.g., Glanzer & Ehrenreich, 1979; Whaley, 1978). Consequently, a major goal of theory has been to explain the obvious importance of word frequency in this and other processes (as recently reviewed by Glanzer & Ehrenreich, 1979).

The distinction between open-class and closed-class words may be of analogously pervasive experimental and theoretical importance for linguistics and psycholinguistics. It apparently reflects differences between syntactic and semantic processing both in normal subjects (e.g., Garrett, 1979) and in aphasics (Caramazza & Zurif, 1976; Zurif & Caramazza, 1976). The "open class" consists of words which bear semantic meaning: nouns, verbs, adjectives, and the like (Garrett, 1979). "Closed-class" words are a subset of the grammatical operators of the language, such as the prepositions, conjunctions, and auxiliaries (Garrett, 1979). (The basis for this labeling is that the number of nouns, verbs, and adjectives is potentially unlimited, while there are no more than 200–400 grammatical words in English.) It might seem that the issues raised by the open/closed class distinction would be independent of those raised by the word frequency effects. Therefore, the reports of Bradley and her colleagues (Bradley, 1978; Bradley, Garrett, & Zurif, 1980) contrasting an almost negligible frequency sensitivity for closed-class words in a lexical decision task with the expected (and reconfirmed) frequency effects for open-class words, have both been unexpected and of potentially great importance for several fields. These results, along with other experiments these investigators reported, have already elicited much comment and speculation (cf. Berndt & Caramazza, 1980; Caramazza & Berndt, in press; Frazier, 1979; Friederici & Schoenle, 1980; Garrett, 1979; Kean, 1980a, 1980b; Swinney, Zurif, & Cutler, 1980; Zurif, 1980). We have therefore attempted to replicate the lexical decision experiments of Bradley et al. with normal subjects. Since our own experiments have been modeled after theirs, we will first summarize their experimental procedures and results in some detail.

Bradley's Experiments

In Bradley's (1978) Experiment 1, a "yes–no" lexical decision was required for 120 words (almost evenly divided between open and closed class) mixed with 80 legal nonwords. Subjects were 30 Massachusetts Institute of Technology undergraduates. Their instructions emphasized response accuracy, but they were otherwise to respond as rapidly as possible. They were warned that the nonwords might look like English words. They were practiced on a block of 30 trials before beginning the two experimental blocks of 4 warm-up and 100 test items each. Stimuli were continuously presented on film as white-on-gray lowercase letters, at intervals of 4 sec (0.5 sec for fixation, 0.5-sec blank field, 1 sec for letter string, 2 sec blank). Reaction times were measured from stimulus onset.

Only the correct "word" responses were analyzed. These reaction time data were further culled to exclude data points that exceeded a two-standard deviation maximum criterion for both the item and the subject. Words with error rates of more than 20% were excluded from further analysis.

We have summarized Bradley's (1978) results in Table 1. She computed a logarithm (base 10) of each word's net Kucera and Francis (1967) frequency per million (adjusted), summing over derivational forms, including all regular syntactic inflections (cf. Stanners, Neiser, Hernon, & Hall, 1979). She then examined the correlation coefficients of item mean reaction times with these log frequencies; Bradley (1977, cited in Bradley, 1978) and others (e.g., Scarborough, Cortese, & Scarborough, 1977) have shown that a logarithmic transformation of the frequency can give a good linear fit to response latencies for the usual range of word frequencies.

In Bradley's (1978) first experiment, the correlation of reaction time with log frequency over the entire frequency range was greater for open- than closed-class words ($r = -0.58$ vs. -0.22). This was even clearer when the confounding length variable was partialed out (partialed $r = -0.54$ for open class vs. -0.03 for closed class). Although the effective correlation for the closed class was nearly zero, Bradley was sensitive to the possibility that reaction time might be a nonlinear function of log frequency (Bradley, 1978, pp. 29, 30), even though she found no significant deviation from linearity in the open class data (Bradley, 1978, pp. 29, 30). She therefore attempted to buttress the whole-range findings by examining the more limited log frequency range of 1.7–3.4, where the two classes overlapped. Here she found a trend in the same direction for partialed reaction times (partialed $r = -0.54$ for open class vs. -0.29 for closed class), but this difference did not reach significance. Even so, she subsequently found further support for the reliability of the whole-range findings in two subsequent experiments. One of these was part of an investigation of nonword interference effects (Experiment 4′, Bradley, 1978, p. 50). In this experiment, she found the whole-range open-class correlation to be -0.72, with closed-class $r = +0.01$. The other experiment was almost identical to her Experiment 1, with the same stimuli, on five hospitalized nonaphasic control subjects as part of her Experiment 6 (Bradley, 1978). In this experiment, the whole range correlation was -0.37 for the open class items, but only $+0.05$ for the closed-class ones.

In a second experiment, Bradley attempted to validate the closed-class frequency insensitivity she had found in her first experiment with a somewhat different set of words. The 60 open-class and 40 closed-class items she used included words selected to give an equal representation of the two classes over the frequency range of overlap (2.1 to 3.5). These words were all monosyllabic, ranging in length from two to five letters. They were presented as in her Experiment 1, along with 80 nonwords, to a different group of 20 subjects. (Whether some of the nonwords had also been used previously was not stated.)

Bradley's (1978) Experiment 2 results are presented in Table 1. As in her Experiment 1, frequency had a significant effect on reaction times for the whole open class ($r = -0.75$), while the correlation for the whole closed class was very small ($r = -0.05$). Additionally, this pattern now remained when the common frequency range (2.1–3.5) was examined: for the open class, $r = -0.55$, which was significantly different from the closed class $r = +0.14$. Furthermore, when we compared the mean reaction times for items of the two classes in the highest frequency portion of this common range (>2.6), the mean open-class reaction time was significantly faster than the closed class's: mean open-class RT = 460 msec, mean closed-class RT = 479, $F(1, 51) = 9.1$, $p = 0.004$ (our calculations). The same trend was present in Bradley's first experiment, although the difference was not significant: mean open-class RT = 471 msec, mean closed-class RT = 476 msec, $F(1, 54) = 0.94$, $p = 0.34$ (our calculations).

If, as Bradley's (1978) results suggest, recognition of a class of items (in a lexical decision task) can be effectively independent of the influence of frequency, then there are profound implications for theories of normal lexical access. In particular, models which deduce the normal frequency effect as a consequence of frequency-ordered serial searches (e.g., Becker, 1976; Rubenstein, Garfield, & Millikan, 1970) or which claim it reflects the "number" of separate internal representations generated by individual word exposure (e.g., Landauer, 1975) will have serious difficulties. They must demand either unjustifiably rapid search rates or entirely separate open- and closed-class access routes to account for the closed-class results. Logogen-like models (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977) can be modified to give class-specific frequency independence a coherent interpretation. For example, one possible mechanism could be to have class-specific adjustments of the logogen response thresholds made after recognition and categorization, which would then affect subsequent recognition. But then the remaining question of why this modification should be made so early in lexical processing—perhaps at the graphemic stage—is left unanswered, and no explanation seems sufficiently cogent.

The implications for both normal subjects and aphasics are even more complicated because of the results Bradley (1978) reported on five patients with Broca's aphasia. They were given what was essentially her Experiment 1. For these patients, in addition to the anticipated open-class frequency sensitivity ($r = -0.37$), the closed class was now equally frequency sensitive ($r = -0.38$).

Bradley (1978) interpreted the results of these and other experiments as evidence for the existence of two separate lexical access systems in normal subjects. One, a general-purpose system, mediates frequency-sensitive recognition of both open- and closed-class words. The other system is specialized exclusively for the closed class. If we also try to account for the faster open-class reaction times found in normal subjects, which apparently equalizes in Broca's aphasia, then we also must postulate that the closed-class recognition mechanism normally inhibits either the operation or the responses of the general-purpose one (Gordon, 1981). Bradley et al. (1980, p. 283) hint at just such an inhibitory action, although on the basis of Bradley's other experiments; Bradley (1978) never raised the issue of the difference in reaction times. She suggested that in Broca's aphasics, the closed-class access system was no longer operational, forcing these patients to use their remaining, class-independent, frequency-sensitive lexical access system.

The next conjecture which might develop from this hypothesis is whether the agrammatism of some patients with Broca's aphasia could be explained simply by damage to the specialized lexical access and recognition system itself, operating relatively peripherally in linguistic processing (Bradley, 1978; Bradley et al., 1980; Garrett, 1979; Zurif, 1980). If so, then it would not be necessary to invoke a more central syntactic deficit to explain agrammatism in these patients.

We wish to emphasize that the lexical decision data have not been the only basis for the interpretations Bradley and her co-workers have advanced about class-specific processing and agrammatism. However, their intepretations of the experiments we have reviewed have certainly lent very important support to their hypotheses. Furthermore, their lexical decision results could impose some very specific requirements on the nature of the two purported recognition mechanisms and their interactions (Bradley, 1978; Bradley et al., 1980; Gordon, 1981).

We therefore felt it was necessary to further investigate the issue of closed-class frequency sensitivity in lexical decision. In particular, Bradley's lexical decision analysis and the theoretical issues posed three interrelated questions which needed to be critically examined:

(1) Whether closed class lexical decision is truly frequency insensitive.

(2) Whether some open class decisions are truly faster than those of closed class items of equivalent or higher frequency. Theoretically, such a finding would have considerable importance, as we have discussed. Practically, confirming such a difference would make it difficult to explain a closed-class frequency insensitivity as the trivial expression of a reaction-time floor effect for this frequency range.

(3) And last, whether the relationship between reaction time and log frequency can be accurately described by a linear function. If so, this justifies linear statistical analysis of the lexical decision data in general.

It might perhaps further constrain theories of frequency sensitivity or insensitivity, as well. If the relationship is *not* linear, and possibly flat at very high frequencies (cf. A. Wingfield, cited by Bradley, 1978, p. 29), then, as Bradley (1978, pp. 29, 30) has pointed out, it would not be surprising to find class differences in frequency sensitivity, since there are no open-class items in the highest frequency ranges.

EXPERIMENTS 1 AND 2

These experiments were nearly identical in design, both patterned after a combination of Bradley's (1978) Experiments 1 and 2. The open- and closed-class words tested included almost all of her words as a subset. One major difference, however, was in our response methodology, for the following reasons. The usual 2-choice, "yes/no" lexical decision task that Bradley (1978) used may actually demand *two* decisions from the subject: first, making the lexical decision itself, and then deciding what response with which hand is required for each category (cf. Egeth, cited and discussed by Pachella, 1974, pp. 48–49). Since our interest was in the "yes–word" decision alone, it seemed reasonable to make the mechanics of response selection as simple as possible. So subjects were asked to make overt responses only for the "yes" decisions; they were not to make any response for "no" decisions. We hoped that this methodologic change would minimize response confusions and errors and reduce the variance of the correct responses. It is possible, of course, that this single response technique would encourage subjects to use a different type of decision process than they would have otherwise used (Pachella, 1974). But, even if their decision strategies did change with the single response technique, whether or not we found class-specific frequency insensitivity would still be relevant to the issues Bradley et al. raise.

Materials and Methods (Experiment 1)

Stimuli and experimental design. A total of 116[1] closed-class and 120 open-class words, including most of the words used in Bradley's first two experiments, were presented (stimuli available upon request). Words with ambiguous class membership such as "can" were excluded.

Each class's word pool was divided into two sets of about 60 words each, for presentation during either the first or the second half of the experiment. In general, words in the first half were taken from Bradley's Experiment 1 and, so, had the same wide variation in length and number of syllables. Words in the second half were taken, if possible, from

[1] The apparent numerical imbalance was the result of some subjects being inadvertently given an earlier version of the final test lists. This led to some of the closed-class words being presented twice, duplicated across the two trial blocks, in place of the desired words. None of the affected subjects ever commented on the recurrences. Although the design was to have been more balanced (120 closed-class words instead of 116), all subsequent discussions of design make explicit allowances for this problem. All results with repeated items were excluded from any data analysis.

her Experiment 2, with a much narrower range of lengths. Four sets of 60 legal, pronounceable nonwords were matched with each subset of word stimuli for length, number of syllables, and initial letter or letter combination, if possible (e.g., "throw"–"thack"). Where feasible, these nonwords were the same as used by Bradley in her Experiment 1.

These sources were mixed into four pairs of blocked presentation lists, with each member of a pair containing about 240 trials exactly divided between words and nonwords. The order of item presentation met the following constraints:

(1) Sequence of item presentation type was pseudorandom, with no more than three sequential items of similar type (open-class word, closed-class word, open-class associated nonword, or open-class associated nonword).

(2) From a master randomized sequence of 240 trials per trial block, three other permutations were created by interchanging (a) word/nonword items, (b) open- and closed-class items, and (c) both (a) and (b).

(3) Any given subject was pseudorandomly given one of the four permutations of the first trial block, followed by a different kind of permutation of the second trial block. The next subject was randomly given one of the three remaining permutations of the first set, and one of the three from the second set (as long as it was different from the first set's permutation type), and so on. This selection cycle was repeated every four subjects.

Subjects. Twenty-eight students of the Johns Hopkins University were paid $3.00 for participation in the experiment. All were native speakers of English with normal or corrected normal visual acuity. Fifteen were female; 24 were right-handed.

Apparatus. Stimuli were presented in lower case letters on a video screen; they subtended no more than 3.0° horizontally and 0.5° vertically at a comfortable reading distance. Responses were by hand-held microswitch using the index finger of the preferred hand. Stimulus presentation and timing were controlled by a PET microcomputer modified for laboratory use.

Procedure. Subjects were instructed orally. They were told that they would be shown letter strings, which they would have to identify as either words (response required) or nonwords (no response required). They were told that they would probably be familiar with all the words, but that the nonwords could look and sound like real words. They were instructed to respond as quickly as possible without making more than occasional mistakes. They were informed that the early items on the first block of trials would be practice (but not how many would be).

Subjects self-paced each trial presentation by a foot-switch. Upon pressing the footpedal, a "*" appeared as fixation point in the center of the screen for 0.5 sec, followed by a blank screen for 0.25 sec, followed in turn by the letter string. The string was displayed for 1.5 sec regardless of whether or not the subject made a response. (As mentioned earlier, Bradley (1978) also displayed stimuli for a fixed time.) Reaction time was measured from stimulus onset (beginning of video sweep). Approximately 0.75 sec after the stimulus had left the screen, the subject was given his/her reaction time (for "yes" responses only) in the upper left corner, as well as the prompt for the next trial in the upper right ("Push foot pedal for next trial"). (Bradley (1978) did not give reaction time feedback after the practice session.) Subjects who were responding slowly were encouraged to go faster. Subjects were free to interrupt the testing at any time to stretch, etc., but most preferred to take their break after the first block of 240 trials. Testing for the whole experiment lasted approximately 45 min.

Data analysis. The first 29 trials of the first block and the first 3 trials of the second block were regarded as practice and warm-up and were excluded from further analysis. Repeated items (if any; see footnote 1) were not analyzed.

Log frequencies per million were calculated using the count of Kucera and Francis (1967). Because it is not clear whether only root forms or whether all derivational forms should be used for assigning frequency, and because our inclusion criteria sometimes appeared to differ from Bradley's (1978), three log frequency values were calculated for each item, in addition to Bradley's own: the log of its root frequency; the log of a summated frequency based on inclusion of all regular derivational forms (this was usually but not always comparable to Bradley's (1978) assignment); and a log frequency based on a "radical" assignment of derivational forms (i.e., including derivations leading to a change in meaning).

Because no item had more than a 7% error response rate, no item was excluded from further analysis.

Reaction time data were analyzed both without modification, and with "extreme" values excluded; one of our extreme value criteria was the same as used by Bradley (1978). Approximately 1.4% (of 6045 usable trials) could be excluded using Bradley's (1978) double criteria (see above). Unlike Bradley (1978), extreme values were simply dropped if this was chosen as the option: if mean reaction times were being used, the item mean reaction time was recalculated without the influence of the presumably erroneous value. If original (per subject) reaction times were to be analyzed, the extreme values were simply not included.

Experiment 2

The results of Experiment 1, which we will discuss, prompted a repetition of the same experimental design, but with words chosen to permit a more exhaustive examination of the very highest-frequency open-class words, and of the lowest-frequency closed-class ones. The 138 open-class words used in our second experiment therefore virtually exhausted the set of possible high-frequency items. The 139 closed-class words used in the second experiment included a number of infrequent but still familiar forms (e.g., "albeit," "thereupon"); words with possibly ambiguous class membership (such as "down" and "inside") were excluded. Most of the nonwords were different from those used in Experiment 1.

Design, apparatus, and procedure. As in our Experiment 1, except that each of the experimental blocks had 280 trials.

Subjects. Twenty students took part. Twelve were female; 17 were right-handed.

Data analysis. No item had more than a 10% error rate, and no item was excluded from analysis. Except for using only mean reaction time data in analysis, the methodology was identical to that for Experiment 1.

Experiments 1 and 2 Results

Our item mean reaction times are plotted in Figs. 1 and 2. (Lists of the items, log frequency assignments, reaction times, error rates, and other data for all our experiments are available from the authors.)

Overall, these mean reaction times are approximately 70–90 msec faster than Bradley's (1978), even with lower error rates (a maximum of 7% in our first experiment and 10% in our second, compared with a maximum error rate of 53% in hers), presumably because of the single-response technique.

Table 1 highlights some of the statistical analyses of our mean reaction time data for both experiments, together with the comparable results for Bradley's Experiments 1 and 2 (taken fron Bradley, 1978). Our tabulated mean reaction time analyses excluded trials using Bradley's criteria, recalculating the means as previously noted. The log of the summated Kucera–Francis frequency, determined by the conservative criteria mentioned earlier, was used in these tabulated calculations.

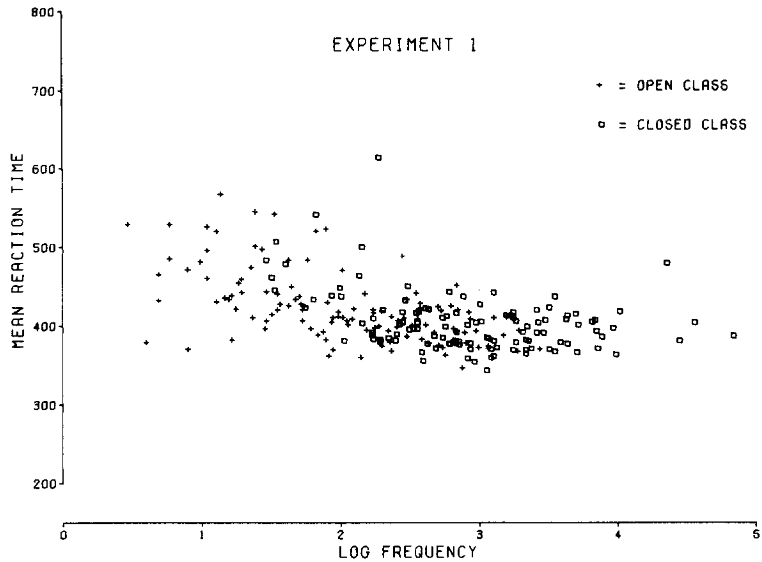

FIG. 1. Mean item reaction times versus logarithm of summated frequency (see text) for Experiment 1 (single-response technique).

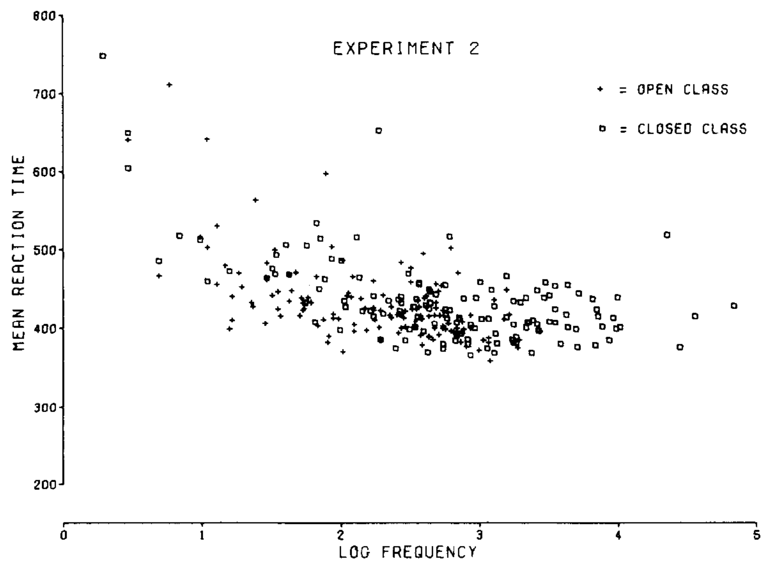

FIG. 2. Mean item reaction times versus logarithm of summated frequency (see text) for Experiment 2 (single-response technique).

TABLE 1
COMPARISON OF EXPERIMENTAL RESULTS—BRADLEY'S (1978) FREQUENCY RANGES

Experiment		Log frequency range[a]		
		Overall	1.7 to 3.4	2.1 to 3.5
Bradley (1978) Experiment 1	Closed class	−0.22 (−0.03)[b]	−0.29 (NSD)	
	Open class	−0.58 (−0.54)	−0.54	
Bradley (1978) Experiment 2	Closed class	−0.03		+0.14
	Open class	−0.75		−0.55
Experiment 1	Closed class	−0.39 (−0.29) (df=114, $p \leq 0.00001$)[c]	−0.44 (−0.36) (df= 81, $p \leq .001$)	−0.30 (−0.30) (df=79, p=0.003)
		z=1.8, p>0.07[d]	z= −0.67, p>0.50	z= −0.18, p>0.84
	Open class	−0.57 (−0.54) (df=118, $p \leq 0.00001$)	−0.35 (−0.36) (df= 77, $p \leq .001$)	−0.27 (−0.26) (df=54, p=0.022)

Experiment 2	Closed class	-0.57 (-0.52) $(df=137, p \leq 0.00001)$	-0.44 (-0.39) $(df=92, p \leq .001)$	-0.28 (-0.27) $(df=84, p=0.005)$
		$z=0.0, p>0.99$	$z=-1.13, p>0.25$	$z=-0.34, p>0.72$
	Open class	-0.57 (-0.58) $(df=136, p \leq 0.00001)$	-0.30 (-0.32) $(df=105, p \leq .001)$	-0.23 (-0.20) $(df=83, p=.019)$
Experiment 3 (two response)	Closed class	-0.43 (-0.33) $(df=134, p \leq 0.00001)$	-0.35 (-0.23) $(df=91, p \leq 0.001)$	-0.19 (-0.08) $(df=83, p=0.04)$
		$z=0.73, p>0.46$	$z=-0.84, p>0.40$	$z=0.47, p>0.62$
	Open class	-0.50 (-0.51) $(df=134, p \leq 0.00001)$	-0.24 (-0.27) $(df=105, p=0.006)$	-0.26 (-0.24) $(df=83, p=0.007)$

[a] As discussed in the text, our log frequency assignments are often slightly different from Bradley's (1978).
[b] Correlation coefficients with length partialed out are in parentheses.
[c] The significance tests are for the simple (unpartialed) correlation coefficients; all are single-tailed.
[d] Two-tailed tests of significant differences between the correlation coefficients.

As can be seen in Table 1, in Experiment 1 our correlation coefficients over the whole frequency range are lower for the closed class than for the open one, as were Bradley's. In our Experiment 2, the closed-class whole-range correlation is either equal to or only slightly lower than the open-class correlation. But this might, of course, be explained in part as an artifact of item selection.

However, our results differ markedly from hers when they are computed over both of the seemingly more appropriate ranges of item frequency overlap that she also used: both classes of items have significant negative correlations of reaction time with log frequency. Moreover, the closed-class correlations are generally more negative than the comparable open-class ones, and are at least statistically indistinguishable. This is true whether or not length is partialed out. Our data were also analyzed in several ways not reported in Table 1: using individual item data (for Experiment 1 only), and using the log frequencies of both the stem words and "radically" assigned summated frequencies, as noted earlier. Also, since correlation analyses are sensitive to the range of items chosen, we also separately analyzed just the subset of items which Bradley (1978) had also used in her experiments (employing her log frequency assignments as well). The results of these different analyses did not deviate substantially from the ones we have tabulated. Our data also do *not* show a difference in the mean reaction times of closed and open class items at the upper end of these frequency ranges (frequencies ≥ 2.6, shown in Table 2).

Our data not only fail to reveal any significant differences between the two classes in their functions of reaction time with log frequency over these ranges, but, more generally, they demonstrate that this function is significantly nonlinear for both classes. Using the individual raw subject data for Experiment 1, we find significant deviations from linearity no matter what class of items is tested: Open class ($F(88, 2998) = 6.1$, $p \leq 0.0001$), closed class ($F(92, 3047) = 4.0$, $p = 0.0002$), or with both sets of items in combination ($F(1158, 6069) = 4.7$, $p \leq 0.0001$), using the statistical tests described by Nie, Hull, Jenkins, Steinbrenner, and Bent (1975, pp. 259–261). With the mean data from Experiment 1, this nonlinearity is only statistically significant with the closed-class items ($F(88, 26) = 1.8$, $p = 0.04$). (Bradley (1978) tested only her open-class mean data for nonlinearity.) In Experiment 2, even with mean data, both classes show a significant nonlinear component: for the open class, $F(100, 36) = 2.07$, $p = 0.0075$; for the closed class, $F(108, 29) = 4.04$, $p \leq 0.0001$; for the combined groups, $F(172, 103) = 2.26$, $p \leq 0.0001$.

By visual inspection of Fig. 1 (and of separate plots of the two classes, not illustrated), this nonlinearity seems to be resolvable into two linear functions: below a log frequency of approximately 2.5, the reaction time distribution might still be described as showing a negative correlation

TABLE 2

COMPARISON OF EXPERIMENTAL RESULTS—LOW- VS. HIGH-FREQUENCY RANGES

		Log frequency range[a]		
Experiment		≤2.5	>2.6	
Bradley's (1978) Experiments 1 and 2 (combined), correlations	Closed class	−0.48 (−0.43)[b] (df=20, p=0.012)[c]	+0.03 (+0.03) (df=69, p=0.42)	Mean RT = 477.3 msec
		z=0.60, p>0.54[d]	z=1.69, p>0.09	
	Open class	−0.59 (−0.53) (df=73, p≤0.001)	−0.31 (−0.29) (df=36, p=0.041)	Mean RT = 466.0 msec
Experiment 1 (single response), correlations	Closed class	−0.43 (−0.44) (df=30, p=0.007)	+0.13 (+0.11) (df=76, p=0.13)	Mean RT = 393.3 msec
		z=0.23, p>0.80	z=1.44, p>0.15	
	Open class	−0.47 (−0.44) (df=82, p≤0.001)	−0.19 (−0.14) (df=28, p=0.15)	Mean RT = 392.5 msec
Experiment 2 (single response), correlations	Closed class	−0.67 (−0.69) (df=44, p≤0.001)	+0.04 (+0.06) (df=84, p=0.34)	Mean RT = 414.7 msec
		z=−0.78, p>0.42	z=2.05, p<0.04	
	Open class	−0.58 (−0.57) (df=76, p≤0.001)	−0.33 (−0.26) (df=45, p=0.012)	Mean RT = 411.7 msec
Experiment 3 (dual response), correlations	Closed class	−0.67 (−0.64) (df=40, p<0.001)	+0.06 (+0.06) (df=84, p=0.28)	Mean RT = 512.1 msec
		z=−1.52, p>0.12	z=2.47, p<0.02	
	Open class	−0.47 (−0.45) (df=75, p≤0.001)	−0.38 (−0.25) (df=45, p=0.005)	Mean RT = 508.0 msec

[a] As discussed in the text, our log frequency assignments are often slightly different from Bradley's (1978) ones. Her assignments were used in recalculating her results over these frequency ranges.

[b] Correlation coefficients with length partialed out are in parentheses.

[c] The significance tests are for the simple (unpartialed) correlation coefficients; all are single-tailed.

[d] Two-tailed tests of significant differences between the correlation coefficients.

with frequency. Above an inflection point at about 2.5–2.6 log frequency, however, the closed class seems to have a relatively flat reaction time distribution, while the open-class reaction times may also stop decreasing. These impressions were subjected to statistical tests as summarized in Table 2, together with our reanalysis of Bradley's (1978) reported data in light of this interpretation.

Using both our data and Bradley's, both classes demonstrate significant and nearly identical sensitivity in the log frequency range below 2.5 (below 316/million Kucera–Francis frequency), with and without partialing out the variance due to length. So, for approximately the lower half of the frequency ranges Bradley (1978) focused on, we find no real difference in frequency sensitivity for the two classes.

In the higher-frequency half of this range, the situation is statistically more complex. As shown in Table 2, the closed class shows no significant effect of frequency for items with log frequencies greater than 2.6 (greater than 399/million, chosen to avoid the inflection point itself) in any of the experiments, whether Bradley's or ours. However, the open-class items do show a frequency sensitivity which reaches statistical significance with Bradley's combined experiments, and in our Experiment 2. However, there are several factors which argue against this finding having theoretical significance. These open-class items have a limited frequency range; the highest open-class log frequency is 3.44. Furthermore, a close examination of Fig. 2 and of the data themselves shows that this "frequency sensitivity" is the product of only 15–25 of the highest-frequency open-class items. And finally, the reaction times of all the open-class items fall within the reaction time borders set by the closed-class items of comparable frequency, and neither of our two experiments reveals any significant differences in mean reaction times for the two classes over this range. Given the limited choice of items possible in this region, the small numbers of items used, and the limitations of frequency assignments in general, we doubt that a strong argument should be made for the behavior of this group of open-class items as being representative of the open class as a whole (as we have sampled it), or as showing a reaction time function which would still have theoretical importance.

EXPERIMENT 3

Our failure to replicate Bradley's (1978) closed-class results in our first two experiments prompted a more exact replication of her experimental methodology, using the standard two-choice, "yes/no" responses. "Yes" responses were made with a microswitch held in the dominant hand; "*no*" responses were made with one held in the other hand. Our Experiment 3 was otherwise identical to Experiment 2; stimuli, instructions, and experimental procedure were the same. Thirty-seven subjects were tested (21 male, 31 right-handed).

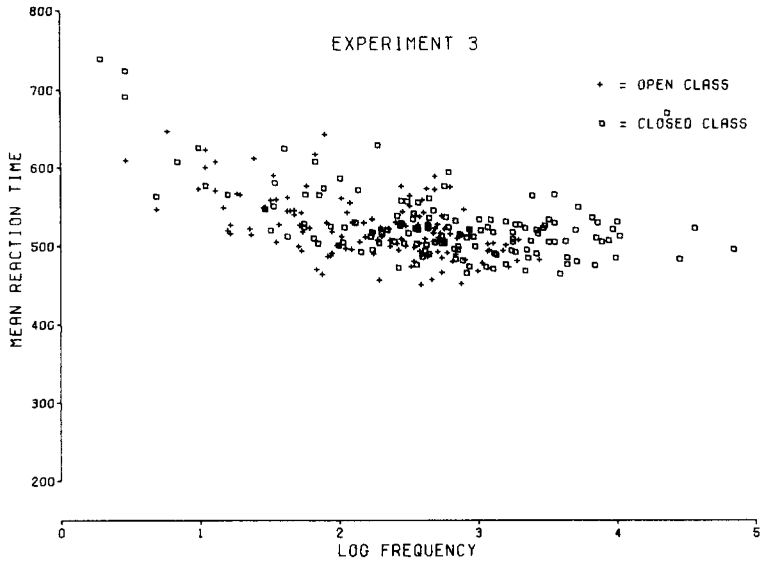

FIG. 3. Mean item reaction times versus logarithm of summated frequency (see text) for Experiment 3 (dual-response technique).

Experiment 3 Results

The data were analyzed as in Experiment 2,[2] and the results are presented in Tables 1 and 2, and plotted in Fig. 3.

Overall, reaction times are about 100 msec slower than those in our previous two experiments (and comparable to or slightly longer than Bradley's); error rates are higher as well, leading to the exclusion of three closed-class items and two open class items from analysis because of error rates in excess of 20% (the same criterion Bradley, 1978, used). The whole-range correlation coefficient for the closed class is slightly less than that of the open class, although this difference is not significant. However, over the log frequency range of 1.7 to 3.4, the simple closed-class correlation coefficient is, if anything, more negative than that of

[2] Additional analyses were also done. Since reaction time has a notoriously non-Gaussian distribution, the simple arithmetic mean is often not a good measure of its central tendency, and the assumptions underlying correlations such as Bradley's (1978) and ours may be unacceptably violated (cf. Wainer & Thissen, 1976). A logarithmic transformation of the individual reaction times is sometimes used to reshape the data into a form more appropriate for statistical treatment (Winer, 1971). Therefore, we also log transformed all of the individual data of Experiment 3 prior to averaging; there were no substantial differences in our results. Since the mean data themselves show a correlation of standard deviation with reaction time, we recalculated the correlations from Bradley's (1978) Experiments 1 and 2 and our own Experiment 3 after log transforming the mean reaction times; again, neither set of results was appreciably altered. We are grateful to an anonymous reviewer for directing our attention to these issues.

open-class; over the 2.1 to 3.5 log frequency range, the closed class still shows significant frequency sensitivity ($r = -0.19$, $df = 83$, $p = 0.04$), which is not statistically distinguishable from that of the open class. (When, as previously, the reaction times of only the words Bradley (1978) also used are analyzed using her log frequency assignments, the same patterns of results over both ranges remain, but the closed-class correlation in the 2.1 to 3.5 range does not quite reach significance: $r = -0.18$, $df = 50$, $p = 0.10$.) Our third experiment therefore also fails to confirm Bradley's (1978) findings, despite using a much more similar experimental procedure.

Moreover, we once again question the validity of these choices of frequency ranges and of the associated statistics, because there is a significant nonlinear component to both the open-class ($F(99, 35) = 1.89$, $p = 0.017$) and overall ($F(174, 96) = 2.03$, $p \leq 0.0001$) data, as well as a nonlinear component for the closed-class data which approaches significance ($F(107, 27) = 1.61$, $p = 0.08$). (With just the set of words Bradley (1978) also used, only the combined data revealed any evidence of a nonlinear component: $F(138, 32) = 1.58$, $p = 0.07$.)

When we then separately examine the log frequency regions on either side of the apparent inflection point, we find that the lower-half, closed-class correlation is markedly negative (−0.67), and not statistically distinguishable from that of the open class (−0.47). And, while the same differences in correlation coefficients are found as before over the highest log frequency range (greater than 2.6), an inspection of Fig. 3 again shows that this is due to the same small set of open-class items that we found were responsible for this result in our first two experiments. The mean reaction times for the two classes over this range are also very close (512 msec for the closed class, 507 msec for the open class), and not statistically different. Again, these data do not seem to be an adequate basis for a distinction between the two classes.

DISCUSSION

If our data are accepted as accurate reflections of the true word frequency effect, they provide several reasons why an analysis such as Bradley's (1978) might have been hazardous and possibly misleading. Since closed-class items are overrepresented in the log frequency region where we see a reaction time floor or saturation, correlations for them done over the entire frequency range will necessarily be lower than those for the open class. It is also hazardous to perform a linear correlation or regression over a range which straddles the inflection point of a nonlinear function. By inspection, this inflection point for our data occurs at a log frequency of about 2.5; the midpoints of Bradley's (1978) two restricted ranges were 2.4 and 2.55, for her Experiments 1 and 2, respectively. Furthermore, since only a small number of items from the two classes share similar frequency ranges, any attempt to ensure an equal balance by frequency risks a biased selection of other item characteristics. Finally, all of these problems may be accentuated by the difficulties of accurately assigning frequencies to such common items, given the precision which the correlational approach demands of the frequency data.

Despite all these caveats, we must conclude that there is no appreciable difference between the frequency sensitivity of the two classes' lexical decision latencies, no matter what the response measure, over the two ranges where they might be appropriately compared: over the lower-frequency range, both show appreciable frequency dependence; over the high range, both have similar reaction time distributions, with the closed class having enough of a range of frequencies to reveal what may be a reaction time floor or saturation effect.

There undoubtedly may be some differences found between the two classes on the lexical decision task; the open-class high-frequency behavior is possibly one of them. We are presently investigating the two classes' frequency sensitivity (or lack of sensitivity) with other methods. However, the present closed-class lexical decision data with normal subjects cannot be advanced as support for class-specific lexical access mechanisms, whatever the other support for this hypothesis. Moreover, in light of these results, the lexical decision performance of the Broca's patients reported by Bradley (1978) cannot be regarded as anomalous. So, whatever impaired mechanisms are responsible for these patients' agrammatic production and comprehension, it does not seem likely to be an impairment in the closed-class lexical access mechanisms which are probed by the lexical decision task.

REFERENCES

Becker, C. A. 1976. Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 556–566.

Becker, C. A., & Killion, T. H. 1977. Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 389–401.

Berndt, R. S., & Caramazza, A. 1980. A redefinition of the syndrome of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics*, **1**, 225–278.

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

Bradley, D. C., Garrett, M. E., & Zurif, E. B. 1980. Syntactic deficits in Broca's aphasia. In D. Caplan (Ed.), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Caramazza, A., & Berndt, R. S. In press. A psycholinguistic assessment of adult aphasia. In S. Rosenberg (Ed.), *Handbook of applied psycholinguistics*.

Caramazza, A., & Zurif, E. B. 1976. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, **3**, 572–582.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. 1977. Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance*. New York: Academic Press. Vol. VI.

Frazier, L. 1979. Constraining models in psycholinguistics. *The Behavioral and Brain Sciences*, **2**, 463–464.

Friederici, A. D., & Schoenle, P. W. 1980. Computational dissociation of two vocabulary types: Evidence from aphasia. *Neuropsychologica*, **18**, 11–20.

Garrett, M. E. 1979. Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production*. London: Academic Press.

Glanzer, M., & Ehrenreich, S. L. 1979. Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, **18**, 381–398.

Gordon, B. 1981. *Lexical access and lexical decision: Mechanisms of frequency sensitivity*. Unpublished doctoral dissertation, Johns Hopkins University.

Kean, M.-L. 1980a. Grammatical representation and the description of language processing. In D. Caplan (Ed.), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Kean, M.-L. 1980b. A note on Kolk's "Judgement of sentence structure in Broca's aphasia." *Neuropsychologia*, **18**, 357–360.

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

Landauer, T. K. 1975. Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology*, **7**, 495–531.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. 1975. *SPSS: Statistical package for the social sciences*. New York: McGraw–Hill. 2nd ed.

Pachella, R. G. 1974. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, MD: Erlbaum.

Rubenstein, H., Garfield, L., & Millikan, J. 1970. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, **9**, 487–494.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. 1977. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 1–17.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. 1979. Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, **18**, 399–412.

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

Wainer, H., & Thissen, D. 1976. Three steps towards robust regression. *Psychometrika*, **41**, 9–34.

Whaley, C. P. 1978. Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, **17**, 143–154.

Winer, B. J. 1971. *Statistical principles in experimental design*. New York: McGraw–Hill.

Zurif, E. B. 1980. Language mechanisms: A neuropsychological perspective. *American Scientist*, **68**, 305–311.

Zurif, E. B., & Caramazza, A. 1976. Psycholinguistic structures in aphasia: Studies in syntax and semantics. In H. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics*. New York: Academic Press.