Bulletin of the Psychonomic Society
1985, 23 (4), 355-358

Recall evidence for a frequency ordered lexicon

PAUL W. FOOS, SUSAN LERO, and PAT McDONNELL
Florida International University, North Miami, Florida

and

MARK A. SABOL
Lynchburg College, Lynchburg, Virginia

Students were asked to recall high-frequency words from the lexicon by listing the 60 most frequent words (Form 10/10; 83 students) or 60 of the 600 most frequent words (Form 10/100; 83 students). The frequencies for recalled words were significantly correlated with the Kučera and Francis (1967) objective count indicating that the lexicon contains frequency information and that the objective count is as useful for recall studies as it is for recognition studies. Differences in performance between Forms 10/10 and 10/100 strongly support the notion that the lexicon is frequency, rather than randomly, ordered. Finally, failure to find significant correlations with number of meanings supports the notion that number of meanings and word frequency are independent properties exerting independent influences on performance. Future research is suggested.

"Lexicon" is the name given to the internal representation of the words in an individual's language. It may contain information on a word's phonological code, morphological structure, grammatical category(ies), meaning(s), and frequency (Foss & Hakes, 1978). We examine several issues concerned with the organization of word-frequency information in the lexicon.

Evidence on the organization, as well as presence, of frequency information in the lexicon comes primarily from studies employing the lexical decision task. In this task, an individual is presented with a string of letters and must decide, as quickly as possible, whether or not the string is a word. The general finding is that high-frequency words are recognized faster than are low-frequency words (e.g., Forster & Chambers, 1973; Rice & Robinson, 1975; Rubenstein, Garfield, & Millikan, 1970), supporting the position that the lexicon is ordered from high to low frequency (Glanzer & Ehrenreich, 1979; Rubenstein et al., 1970).

One potential problem with the evidence cited for a frequency ordered lexicon is that the lexical decision task is a recognition task, but the lexicon is used in recall (e.g., speech production) as well as in recognition (e.g., speech comprehension). Recall and recognition, however, differ in a number of important ways. For example, retrieval operations appear to be independent (e.g., Flexser & Tulving, 1978) and different (e.g., Anderson & Bower, 1974; Foos & Clark, 1983) for the two tasks. It may, then, be a mistake to posit a frequency ordered lexicon solely on the basis of evidence from a recognition task. The obtained evidence may be a function of the task used rather than a reflection of a frequency ordered lexicon.

The present study provides evidence from a recall task. Individuals were asked to recall the most frequent words in their experience with the English language. A correlation between performance on this task and an objective frequency count (i.e., Kučera & Francis, 1967) would provide further support for a frequency ordered lexicon. Such a correlation, however, could also be taken as support for a randomly ordered lexicon (Landauer, 1975). In a randomly ordered lexicon, frequent words would be easier to find because they have many more representations than infrequent words. To separate these two possible organizational schemes, two forms of the present task were used.

On Form 10/10 of the present task, individuals were asked to recall the 10 most frequent words, followed by the 10 next most frequent words, and so on for a total of 60 words. If the lexicon is either frequency or randomly ordered, then individuals should be fairly successful on this form of the task. On Form 10/100, individuals were asked to recall 10 of the 100 most frequent words, followed by 10 of the next 100 most frequent words, and so on for a total of 60 words. Again, if the lexicon is frequency ordered, individuals should be successful on this form of the task. If the lexicon is randomly ordered, however, then performance on this form should be quite difficult. Ten of the 100 most frequent words could be recalled by selecting the first 10 words encountered in the search. They are likely to be very high-frequency words. In recalling 10 of the next 100 most frequent words, however, one would be less successful by simply selecting the next 10 words encountered. Those words are likely to be among the first 100 most frequent words. To perform well on Form 10/100, one must be able to select

The mailing address for P. W. Foos, S. Lero, and P. McDonnell is: Department of Psychology, Florida International University, Bay Vista Campus, North Miami, FL 33181. M. A. Sabol's mailing address is: Department of Psychology, Lynchburg College, Lynchburg, VA 24501.

frequent words and also reject those which are too frequent. This should be particularly difficult as one proceeds through the form (e.g., the fifth or sixth set of "100 next most frequent words"). If the lexicon is randomly ordered, then individuals receiving Form 10/100 may be unable to comply with the instructions. Their responses would turn out to be essentially the same as those of individuals who received Form 10/10.

Some recent work by Jastrzembski (1981) strongly suggests that the number of meanings that a word has is as important as the word's frequency for performance on a number of tasks, and that number of meanings and frequency are independent properties exerting independent influences on performance. The present study assesses this independence by correlating performance with number of meanings as well as objective frequency. If the two are independent, then in the present study, in which individuals are asked to respond on the basis of frequency, a correlation with objective frequency should occur and a correlation with number of meanings should not.

In sum, the present study examines the recall of high-frequency words from the lexicon. Correlations between performance on Forms 10/10 and 10/100 and an objective frequency count (i.e., Kučera & Francis, 1967) are used to assess the hypothesis of a frequency ordered lexicon (Glanzer & Ehrenreich, 1979; Rubenstein et al., 1970). Correlations with number of meanings are used to assess the independence of word frequency and of number of meanings (Jastrzembski, 1981).

METHOD

Participants

Participants were 166 undergraduate students ranging in age from 18 to 29 years who were randomly divided into two groups. Students voluntarily participated at the end of regular classes at Florida International University and Indiana University of Pennsylvania.

Materials

Students received one of two forms. Form 10/10 asked individuals to list the 10 most frequent words on the first page, the 10 next most frequent words on the second page, and so on for a total of six pages (60 words). Form 10/100 asked individuals to list 10 of the 100 most frequent words on the first page, 10 of the next 100 on the second page, and so on for a total of six pages (60 words). Each of these forms was given to 83 participants. On the last page, students estimated the number of books, textbooks, and newspapers that they read and the number of hours that they spend watching television for given time periods.

Procedure

Students were run in groups ranging in size from 7 to 30 individuals. They were told that some words occurred more frequently than others and that the experimenter was interested in finding out what the most frequent words were. After receiving a form, they were told to use their total experience, written and spoken, with words in filling out the form. Previous work (Shapiro, 1969) shows no major differences between estimates based on written versus spoken language. All subjects were given as much time as they needed to complete the forms.

RESULTS AND DISCUSSION

For each form, all words were ranked according to how many participants listed them among their 60 words. Words that were listed by only one participant were excluded from all analyses. Table 1 shows the words and numbers of listing participants. Form 10/10 produced 120 different words that occurred two or more times; Form 10/100 produced 98 such words. A total of 75 words were common to the two forms. Because there were several ties in the ranks, Pearson, rather than Spearman, correlations were computed (Hays, 1963). The rank order of the 75 common words on Form 10/10 was significantly correlated with their rank order on Form 10/100 [$r(73) = .74$, $p < .01$].

For correlations between performance on Forms 10/10 and 10/100 and objective frequency, we used the Kučera and Francis (1967) count. The ranking of words from Form 10/10 (based on the frequencies given in column 2 of Table 1) was significantly correlated with the objective ranking from the Kučera and Francis count [$r(118) = .39$, $p < .01$]. The ranking of words from Form 10/100 (based on the frequencies given in column 3 of Table 1) was also significantly correlated with the objective ranking [$r(96) = .52$, $p < .01$]. Finally, across both forms, the total ranking of words (based on the frequencies given in column 4 of Table 1) was significantly correlated with the objective ranking [$r(141) = .44$, $p < .01$]. These results support the notion of a frequency ordered lexicon. By themselves, these correlations do not, however, rule out the possibility of a randomly ordered lexicon.

To determine whether the lexicon is frequency or randomly ordered, one must examine performance on Form 10/100. If the lexicon is randomly ordered, then the instructions of Form 10/100 should have been very difficult, if not impossible, to follow. That is, one would encounter many high-frequency words while searching the lexicon but would be unable to determine whether those words belonged among the 100 most frequent words, 100 next most frequent words, and so on. Performance on Form 10/100 would then be very much like performance on Form 10/10. If, however, the lexicon is frequency ordered, then performance on the two forms should differ. Individuals receiving From 10/100 would be able to

Table 1
Form 10/10, Form 10/100, Total Recalled Frequencies, and Kučera and Francis' Frequencies and Ranks for 143 Words

	Form		Recall	Kučera & Francis	
Word	10/10	10/100	Total	Frequency	Rank
the	60*	54	114	69,971	1
to	57	48	105	26,149	4
a	46	51	97	23,237	5
you	51	44	95	3,286	32
I	56	37	93	5,173	19
and	50	42	92	28,852	3
is	50	34	84	10,099	8
no	41	37	78	2,201	48
it	43	28	71	8,756	12
yes	35	36	71	144	107
what	30	34	64	1,908	52
he	38	23	61	9,543	10
go	33	27	60	626	74
me	34	25	59	1,181	60

Table 1 (continued)

Word	Form 10/10	Form 10/100	Recall Total	Kučera & Francis Frequency	Kučera & Francis Rank
she	36	22	58	2,859	36
are	34	23	57	4,393	23
why	27	30	57	404	80
this	28	24	52	5,146	20
love	20	31	51	232	92
of	19	29	48	36,411	2
how	24	23	47	834	65
that	25	21	46	10,595	7
was	30	16	46	9,816	9
when	24	21	45	2,331	44
where	25	19	44	938	62
we	22	21	43	2,653	40
work	23	19	42	760	68
they	23	18	41	3,618	29
study	18	23	41	246	89
food	16	25	41	147	106
in	25	14	39	21,341	6
but	22	16	38	4,388	24
an	18	20	38	3,747	28
home	17	20	37	547	76
on	22	14	36	6,742	15
do	17	19	36	1,363	57
class	22	14	36	207	98
drink	20	15	35	82	118
for	21	13	34	9,489	11
can	20	14	34	1,772	55
not	23	10	33	4,609	22
book	15	18	33	193	100
car	15	17	32	274	87
sleep	13	19	32	65	121
good	16	15	31	807	66
eat	17	14	31	61	123
hate	19	11	30	42	127
clothes	13	16	29	89	116
am	18	10	28	228	93
who	13	15	28	2,252	45
time	7	21	28	1,599	56
there	10	17	27	2,724	37
because	18	9	27	883	64
see	17	10	27	772	67
school	14	13	27	492	78
at	17	9	26	5,378	17
come	17	9	26	630	73
girl	16	10	26	220	94
will	15	8	23	2,244	46
if	11	11	22	2,199	49
so	7	15	22	1,984	51
don't	13	9	22	489	79
boy	10	12	22	242	90
maybe	9	13	22	134	109
have	21	0	21	3,941	27
them	6	15	21	1,789	54
shit	21	0	21	2	140.5
wee	20	0	20	3,284	33
with	11	8	19	7,289	13
nice	8	11	19	75	119.5
or	12	6	18	4,207	26
her	7	11	18	3,037	34
my	17	0	17	1,319	58
look	10	7	17	399	81.5
party	0	17	17	216	95.5
dog	16	0	16	75	119.5
beer	16	0	16	34	130.5
one	15	0	15	3,292	31
did	15	0	15	1,044	61

Table 1 (continued)

Word	Form 10/10	Form 10/100	Recall Total	Kučera & Francis Frequency	Kučera & Francis Rank
down	15	0	15	895	63
want	8	7	15	329	85
money	0	15	15	265	88
read	6	9	15	173	101
can't	8	7	15	169	102
him	6	8	14	2,619	41
up	14	0	14	1,895	53
friends	0	14	14	162	103
hi	7	7	14	6	138
be	12	0	12	6,377	16
walk	12	0	12	100	115
room	0	11	11	383	83
hello	0	11	11	10	137
from	10	0	10	4,369	25
sexy	10	0	10	2	140.5
out	9	0	9	2,096	50
now	0	9	9	1,314	59
bed	0	9	9	127	111
test	9	0	9	119	112.5
smoke	9	0	9	41	128.5
their	8	0	8	2,670	39
going	0	8	8	399	81.5
really	8	0	8	275	86
run	8	0	8	212	97
friend	8	0	8	133	110
please	0	8	8	62	122
shoes	8	0	8	44	126
by	7	0	7	5,305	18
house	7	0	7	591	75
went	7	0	7	507	77
play	7	0	7	200	99
hair	0	7	7	148	105
mail	0	7	7	47	125
shopping	0	7	7	27	132
fuck	0	7	7	4	139
TV	7	0	7	0	143
had	6	0	6	5,133	21
which	6	0	6	3,562	30
has	6	0	6	2,439	43
us	6	0	6	672	71
president	6	0	6	382	84
job	0	6	6	238	91
music	6	0	6	216	95.5
bad	0	6	6	142	108
brothers	0	6	6	41	128.5
pen	6	0	6	18	135
shower	0	6	6	15	136
all	5	0	5	3,001	35
parties	5	0	5	59	124
never	4	0	4	698	69
day	4	0	4	686	70
great	0	4	4	665	72
talk	0	4	4	154	104
window	0	4	4	119	112.5
sun	4	0	4	112	114
sex	0	4	4	84	117
damn	0	4	4	34	130.5
honey	4	0	4	25	133
okay	4	0	4	20	134
showers	0	4	4	3	142
more	3	0	3	2,216	47
his	2	0	2	6,997	14
would	2	0	2	2,714	38
been	2	0	2	2,472	42

**Number of participants listing word.*

choose high-frequency words and reject others as being too frequent (i.e., too high in the ordered lexicon). In sum, the responses of participants who received Form 10/100 should differ from those of participants who received Form 10/10, if the lexicon is frequency ordered. Three lines of evidence show that performance on these two forms did differ.

The first line of evidence is the number of words listed by only one participant. If a group of participants are selecting from 600 (on Form 10/100) rather than 60 (on Form 10/10) top words, one would expect to find a greater number of words listed by only one participant and fewer words listed by more than one participant. This is the case, as the number of words listed by only one participant on Form 10/100 was 2,321 and the number on Form 10/10 was 1,868. As previously stated, the number of words listed by more than one participant was 98 on Form 10/100 and 120 on Form 10/10. These distributions of once-occurring and more than once-occurring words on the two forms were significantly different [$\chi^2(1) = 9.14$, $p < .01$].

The second line of evidence involves comparing the objective (i.e., Kučera & Francis) ranks for the 45 words of Form 10/10 and the 23 words of Form 10/100 which were not common to the two forms. The mean objective rank for these Form 10/100 words should be lower because they would have been selected from an individual's top 600 rather than 60 words. As predicted, the mean rank for these Form 10/100 words (107.52) was significantly lower than the mean rank for these Form 10/10 words (79.81) [$t(66) = 3.03$, $p < .01$].

The third line of evidence is the proportion of words common to the two forms for each of their six pages. On page 1, Form 10/10 asks for the 10 most frequent words while Form 10/100 asks for 10 of the 100 most frequent words. On this page, one might expect a fair proportion of the words listed to be common to the two forms. On page 2, Form 10/10 asks for the 10 next most frequent words, and From 10/100 asks for 10 of the 100 next most frequent words. If participants were able to follow the instructions of Form 10/100, one would expect to find a smaller proportion of common words on this page. In fact, if participants were able to follow the instructions of both forms perfectly, one would expect to find common words only on the first page of each form. At the very least, the proportion of words common to the two forms should decline as one continues from the first through the sixth page. This is exactly the case as the proportions of common words for pages 1 through 6 were .62, .47, .06, .11, .10, and .03, respectively. Together these three lines of evidence argue strongly against a randomly ordered lexicon and support the hypothesis of a frequency ordered lexicon.

To assess the independence of the properties, word frequency and number of meanings, the frequencies for words on Forms 10/10 and 10/100 were correlated with numbers of meanings for those words. Number of meanings were determined by consulting *The American Heritage Dictionary of the English Language* (1976). The correlations between number of meanings and the frequencies from Form 10/10 (column 2 of Table 1), Form 10/100 (column 3 of Table 1), and total frequency (column 4 of Table 1) were not significant [$r(118) = -.18$, $r(96) = -.15$, and $r(141) = -.07$, respectively]. Because all correlations with objective frequencies were significant and none with number of meanings reached significance, the present results support Jastrzembski's (1981) position that word frequency and number of meanings are independent properties exerting independent influence.

In sum, the present study provides evidence from a recall task concerning the organization of the lexicon. The vast majority of previous studies have used recognition tasks that are in many ways different from recall tasks. It is, thus, reassuring to find that the same kind of results can be obtained from a recall task. Individuals appear to have had little difficulty in naming high-frequency words on either of the forms used. Furthermore, differences in performance between Forms 10/10 and 10/100 make the hypothesis of a randomly ordered lexicon (Landauer, 1975) extremely unlikely. Only some form of frequency ordered lexicon (e.g., Glanzer & Ehrenreich, 1979; Rubenstein et al., 1970) can easily accommodate the present and past findings.

REFERENCES

ANDERSON, J. R., & BOWER, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, **2**, 406-412.

FLEXSER, A. J., & TULVING, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, **85**, 153-171.

FOOS, P. W., & CLARK, M. C. (1983). Learning from text: Effects of input order and expected test. *Human Learning*, **2**, 177-185.

FORSTER, K. I., & CHAMBERS, S. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, **12**, 627-635.

FOSS, D. J., & HAKES, D. T. (1978). *Psycholinguistics*. Englewood Cliffs, NJ: Prentice-Hall.

GLANZER, M., & EHRENREICH, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **18**, 381-398.

HAYS, W. L. (1963). *Statistics*. New York: Holt, Rinehart, & Winston.

JASTRZEMBSKI, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, **13**, 278-305.

KUČERA & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

RICE, G. A., & ROBINSON, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, **3**, 513-518.

RUBENSTEIN, H., GARFIELD, L., & MILLIKAN, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **9**, 487-494.

SHAPIRO, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning & Verbal Behavior*, **8**, 248-251.

The American heritage dictionary of the English language. (1976). Boston: Houghton Mifflin.

(Manuscript received for publication March 18, 1985.)