JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 9, 487–494 (1970)

Homographic Entries in the Internal Lexicon[1]

HERBERT RUBENSTEIN
Lehigh University, Bethlehem, Pennsylvania 18015

LONNIE GARFIELD
The Wright Institute, Berkeley, California

AND

JANE A. MILLIKAN
McLean Hospital, Belmont, Massachusetts

The task was to distinguish between English and nonsense words, which were displayed singly. The display persisted until *S* pressed the *yes*-key if he thought the stimulus was English or the *no*-key if he thought it was nonsense. The response times were faster for English than nonsense, faster for English words of higher frequency than lower frequency, and faster for homographs than nonhomographs. It is hypothesized that word recognition in general requires consulting the internal lexicon. A model of the underlying processes is sketched which proposes that words of higher frequency are recognized sooner because their lexical entries are marked earlier for comparison against the stimulus information. It is also proposed that homographs are recognized sooner than nonhomographs since homographs have more lexical entries available for comparison against the stimulus information.

The results of a pilot experiment carried out by Thomas G. Bever, H. Rubenstein, and Roberta Kelly indicated that the visual recognition threshold was lower for homographs, words with more than one meaning like *yard*, *still*, *stain*, than for nonhomographs of the same length and frequency of usage. This result was interpreted as indicating that the process of recognizing a word requires that *S* test his hypothesis against the information about English words stored in his long-term memory, in his internal lexicon. How else could *S* know whether what he thought he saw and intended to report was indeed an English word? The particular finding that the threshold was lower for homographs than for nonhomographs suggested, furthermore, that there were more entries for a homograph in the internal lexicon, perhaps a different entry or set of entries for each different meaning of the homograph.

The experiment described in the present paper was designed primarily to test this view that the recognition of words involves consulting the internal lexicon. If it is true that this consultation is necessary in a recognition task to confirm for the *S* that his hypothesis is, in fact, an English word, then it certainly should be necessary in an experiment in which the task is precisely that of discriminating between English and nonsense words. And the task in our experiment here was just this: to indicate whether a displayed word was English or not, *without being required to identify it*. If, in this simple task, a significant difference in reaction time should be obtained between various kinds of English words, for example,

[1] The data were gathered at the Center for Cognitive Studies, Harvard University, with the support of ARPA Project SD-187 and processed at Lehigh University with the support of NSF Grant GS 2001. The authors express their gratitude to Spafford Lewis, Richard Thomas, and Mollie A. Rubenstein for their assistance in the statistical analysis.

homographs versus nonhomographs, we would take this as supporting the notion that word recognition involves lexical lookup.

Even though the task of the *S* was to discriminate between English and nonsense words, our interest was, of course, in the effects of variables incorporated in our set of English words. In addition to homography, we considered two other variables which we hoped might shed some light on the structure of the internal lexicon: word frequency and concreteness of meaning. We selected these variables because it seemed reasonable that word frequency, homography, and concreteness of meaning come into play at successively deeper stages in the recognition process. Furthermore, if concreteness should turn out to be a significant factor, this would be confirmation of the interesting hypothesis that recognition is not merely a matter of locating an entry but also of retrieving the semantic information bonded to it.

METHOD

Procedure

The *S*s were told that they were going to see isolated words, some of which were English and some of which were nonsense but looked like English, that is, nonsense words following English orthographic and phonological rules. They were instructed to press the *yes*-key or the *no*-key as soon as they decided that the word was English or not. The word was erased when either key was pressed.

The Lexigraph program developed by Dr. Daniel Forsyth at the Harvard Center for Cognitive Studies was used to present the words on the scope of a PDP-4 computer and to measure and store the time elapsed from the stimulus presentation to *S*'s pressing a key. After *S* pressed the key, 2 sec passed before the presentation of the next word. Forty-one practice words (half English and half nonsense) were first presented to the *S* to familiarize him with the task and the equipment. The *S*s were tested individually in a single session lasting about an hour. The *S*s were 39 Harvard-Radcliffe undergraduates, paid for their participation.

Materials

As was stated above, the nonsense words, all 4–5 letters long, were constructed according to the rules of English orthography and phonology, for example, *teft, hosk, gloan.* No effort was made to select letter strings of particular probabilities except to avoid especially rare strings. One hundred and sixty-five different nonsense words were presented. These words were randomly distributed among 180 English words likewise 4–5 letters long. These words fall into the subsets shown in Table 1.

The word frequencies, the number of occurrences in 4.5 million words of magazine text, were taken from the Lorge Magazine Count.[2] The mean frequencies for a given frequency range varied relatively little over the other conditions (see Table 1). Note that for homographs the word frequency is the *sum* of the frequencies of usage of the word-form in each of its meanings.

Most of the homographs used had two distinct meanings. There were a number, however, with three meanings, and several with four meanings. We avoided homographs in which the difference in meaning was primarily a reflection of the difference in syntactic class; however, two such words were used: *place, water* (noun and verb).

All stimulus words were pretested for homography in this manner: A group of 20 *S*s, not used in the present experiment, was given the list of words used in this experiment and asked to write down the first meaning that came to mind for each word. On the basis of an intuitive notion that the homograph effect would be greatest when the meanings were equiprobable, we limited our selection as far as possible to those words in which both meanings were frequent. On the average the frequencies of the meanings were in the ratio 13:7. The English words are listed in the Appendix.

The decision whether a meaning of a word was concrete or abstract was made by the authors. We considered a meaning concrete only if its referent was perceptible by the senses.

RESULTS

The accuracy of discrimination was high under all conditions, 92% or better. The accuracy data are given in Table 1.

As is usual with reaction times (RTs) to verbal stimuli, there was great variance even

[2] We used a computer printout of unknown origin, which had separate entries for the various inflected forms of nouns and verbs in contrast to the published version (Thorndike & Lorge, 1944), which has a single entry for such forms as *walk, walks, walked, walking.* The frequencies according to the published version were slightly larger on the average for the homographs than for the nonhomographs but not enough to account for the obtained differences: nonhomographs, concrete, 32, 233, 2128; abstract, 25, 209, 1879; homographs, concrete, 49, 230, 2357; concrete/abstract, 31, 289, 2027.

TABLE 1

MATERIALS AND RESULTS[a]

		Freq.	N	Lgth.	Fam.	Cor.	RT
Nonhomograph	Concrete	19	20	4.5	58	95	899
		174	20	4.5	42	97	802
		1676	20	4.6	19	99	740
	Abstract	21	20	4.5	59	92	941
		171	20	4.5	41	97	793
		1705	20	4.4	21	98	741
Homograph	Concrete	22	14	4.4	58	95	889
		173	10	4.3	45	98	787
		1905	4	5.0	21	100	725
	Conc./Abst.	16	6	4.3	62	95	861
		182	10	4.6	44	95	769
		1686	16	4.2	25	98	712
Nonsense			165	4.5		96	1021

[a] Freq. = Mean word frequency according to special form of Lorge Magazine Count. N = Number of stimulus words. Lgth. = Mean word length in letters. Fam. = Rank ordering of words in all-over familiarity from 1 (most familiar) to 75. Cor. = Per cent correct responses. RT = Antilogarithm of the mean of the logarithms of (correct) response times in msecs (mean of *S*s' means).

within a *S*. To reduce the effect of deviant RTs, all the class means reported below were obtained by averaging the *S*s' means of logarithmic transforms of original RTs and then taking the antilogarithm of this average.

As might be expected, *S*s were faster in deciding that a word was English than that it was nonsense. Thus the longest RT for any English condition, the lowest frequency class of abstract nonhomographs, was 941 msec as compared to the mean RT of 1021 msec for nonsense words (statistically significant, $p < .01$, according to two-tailed sign test). The results for the various English conditions are shown in Table 1. The RTs for the individual English words are given in the Appendix.

An analysis of variance performed on the data showed that word frequency and homography had statistically significant effects: Frequency, $F(2, 38) = 45.53, p < .001$; Homography, $F(1, 38) = 10.72, p < .005$. Concreteness was significant within homographs, $F(1, 38) = 17.77, p < .001$; this was the only significant interaction.

We obtained substantially the same homograph effect in two earlier versions of the present experiment described in the *Annual Report of the Center for Cognitive Studies, Harvard University*, for 1966–1967, pp. 25–26. However, this agreement does not rule out the possibility of the effect of some unknown variable in the stimulus materials since almost the same set of homographs was used in all three experiments. Before we got too deeply involved in trying to explain an effect that did not exist, we considered the possibility that our results were due to the inadequacy of our frequency count as a measure of familiarity.

It may be that the recognition of visually presented words is affected not only by the number of times *S*s have seen the words in printed matter but also by the frequency with which they have used the word in writing or speech or have heard the word. We thought, therefore, that we should check to see whether homographs and nonhomographs were equally familiar. Accordingly Spafford Lewis of Lehigh University undertook to determine the familiarity of the words used in the present experiment. He had 30 *S*s (Lehigh undergraduates) rank-order the English words in sets of 75 (the meanings were not given) according to overall familiarity which was defined for the *S*s as "how often you use, see, or hear a given word relative to the other words." The mean familiarity rating for each word class is given in Table 1. As can be seen the mean familiarity of corresponding homographs and nonhomographs is almost identical. The correlations between familiarity and the logarithm of word frequency turned out to be quite high: for all words $r = -.88$; specifically, for nonhomographs $r = -.87$; for homographs $r = -.91$. (The correlations are negative since the greater the familiarity, the lower was the rank number.) The reliability of the *S*s' judgments was very high. For 10 *S*s who were asked to rerank the words after a space of 2 or 3 weeks, the correlation between the first and second rankings came to $r = .98$. We conclude that the lower RT for homographs cannot be attributed to greater familiarity.

Our data also gave some positive evidence that it was the multiple meanings that were responsible for the lower RT of homographs in that we found that homographs with more than two meanings had shorter RTs on the average than homographs with just two meanings. When we partitioned all the homograph stimuli into those with two meanings and those with three or four (we could not do this with Frequency 100 since there were too few of these with more than two meanings), we found that those homographs with two meanings had significantly longer RTs than those with more than two: Frequency 10, 909 msec (two meanings, $N = 14$) versus 824 msec (more than two meanings, $N = 6$); Frequency 1000, 718 msec (two meanings, $N = 8$) versus 713 msec (more than two meanings, $N = 12$); $F(1, 38) = 17.09$, $p < .001$. The effect of the number of meanings is significantly greater in the lower frequency than in the higher frequency homographs, $F(1, 38) = 8.52$, $p < .01$.

Discussion

The main findings of the experiment emerge clearly: (*a*) The RT is shorter as word frequency is greater. Furthermore, the effect of word frequency, about 75 msec per logarithmic unit of frequency, is roughly the same in homographs and nonhomographs. (*b*) The RT is shorter as the number of meanings is greater. Thus, not only is the RT for homographs shorter than the RT for nonhomographs but the RT for homographs with three meanings is shorter than the RT for homographs with two meanings. (*c*) Homographs with one meaning concrete and the other abstract have shorter RTs than homographs with both meanings concrete. (*d*) The RT for English is shorter than the RT for nonsense.

We shall first consider the results relating to the concreteness variable. Other investigators have reported finding that words with concrete meanings have a lower recognition threshold than words with abstract meanings, but they did not control for homography: Riegel and Riegel (1961) and Spreen, Borkowski, and Benton (1967). Our finding of no difference in RT between concrete and abstract nonhomographs supports Winnick and Kressel (1965), who also reported obtaining no difference in the visual thresholds of concretes and abstracts. However, our finding of a very substantial and statistically significant difference between homographs with only concrete meanings and homographs with both concrete and abstract meanings is puzzling. If there should be any difference, we would have expected that words with only concrete meanings would have the shorter RTs on the hypothesis that meanings which are richer in sensory information are more readily retrieved. No explanation can be found in terms of differences in frequency, familiarity, word length (see Table 1), number or relative frequency of meanings. We considered the hypothesis that the obtained difference in RT was due to a difference in the associative fields of these two kinds of homographs, that is, the entries of homographs with concrete and abstract meanings have more associates bonded to them than the entries of homographs with only concrete meanings. Mollie A.

Rubenstein ran an experiment to determine how many associations could be made to each of these words in a minute. The result was the opposite of what was hypothesized: homographs with only concrete meanings elicited significantly more associates than homographs with abstract and concrete meanings. Number of associations, then, offers no explanation. At any rate, the obtained difference between these two types of homographs differing only in kinds of meanings leaves open the possibility that word recognition involves the retrieval of semantic information. This should not seem too strange. After all we usually go through the process of recognizing a word only to find out its meaning.

Despite the fact that the task set for *S* in the present experiment was that of merely deciding whether a word was English or not, it is likely that *S* identified the particular word presented. Otherwise, we could account neither for the word-frequency effect nor for the very high degree of accuracy in *S*s' judgments. This is not to say that there may not be some very important differences between the processing in the usual word-recognition task where *S* is asked to identify the stimulus and the processing involved in our task. In word-recognition experiments the amount of stimulus information available to the *S* is fixed by duration of presentation, luminance, or signal-to-noise ratio. In our experiment, on the other hand, *S* had the opportunity to extract as much information as he required to reach his criterion for response. We shall be able to evaluate the force of these differences only when we have RT data on the same set of words both when the task is explicitly identification and when the task is nonsense–English discrimination.

The major difficulty in explaining our primary results lies in reconciling the effect of word frequency with the homograph effect, the effect of the number of meanings. When we say that we have matched our stimuli in word frequency, that is, we have chosen homograph *A* and nonhomograph *B* from the same word-frequency range, we are saying that the sum of the frequencies of the meanings of *A* approximate the frequency of *B*. Hence, the frequency of any meaning of *A* is less than the frequency of *B*. Thus, if we claim that the lower RT of *A* is due to the greater number of its lexical entries, we are saying that *S* locates this entry or meaning of *A* more quickly than he locates *B* despite the fact that *B* is more frequent.

There is another frequency effect that we must take into account. Earlier in discussing the selection of materials we said that we felt that whatever the effects of homography, they would be strongest in homographs in which the meanings tended to equiprobability. This intuition has been supported by the results of an experiment carried out last year by Rubenstein and Spafford Lewis (in preparation): homographs in which the meanings are closer to equiprobability are discriminated from nonsense more quickly than those in which the meanings differ greatly in probability of occurrence. This means that the relative frequencies of lexical entries (on the hypothesis that different meanings of a homograph have separate entries) constitute a factor independent of the frequency of the word as a string of letters or phonemes.

The explanatory model that we propose, very tentatively, of course, involves these processes: (*a*) *quantization*, the division of the stimulus into segments and the assignment of these segments to letters (phonemes in the case of an auditory stimulus); (*b*) *marking*, a process in which the output of quantization marks some subset of lexical entries as being in agreement with it; (*c*) *comparison* of subsequent quantization outputs with marked entries; (*d*) *selection* of the marked entry which meets the accuracy criterion set by the *S*. This is a parallel processing model. As soon as part of the stimulus is quantized, those lexical entries that agree with the quantization are marked. By "marking" we mean no more than that certain entries are distinguished in some way to be ready for comparison against subsequent quantization output. (There is feedback. If there is no lexical entry that agrees, then the quantization is checked. And marked entries may serve to guide further quantization.) Subsequent quantization outputs are compared against marked entries. Entries which do not fit the new information in these outputs are eliminated. The comparison process is carried on until only one entry remains which meets *S*'s accuracy criterion (or no entry remains—in which case *S* decides the stimulus is nonsense).

The findings of the experiment reported here are explained as follows: (*a*) The word frequency effect results from the action of the marking process in marking entries of the highest frequency range first. If none of these entries fits the subsequent quantization output, entries of successively lower frequency ranges are marked. For example, if the first quantization output were *ti--*, it would mark entries like *time* and *till*. Only when these are eliminated by the output *tiv-* would the rare word *tivy* (if *S* knew it) be marked. (*b*) The homograph effect results from the scanning procedure in the comparison process, that is, the order in which marked entries are compared against later quantization outputs is random, at least with regard to frequency. Thus, the fact that a homograph has more entries makes it likely that some one of these entries will be taken up for comparison sooner than an entry for a nonhomograph. The finding that homographs with three meanings have shorter RTs than homographs with two meanings supports this view. (*c*) The effect of the relative frequency of the entries of a homograph, that is, the finding that homographs whose meanings tend to equiprobability have shorter RTs than homographs whose meanings differ greatly in probability, results from the frequency ordering in marking. As was pointed out above, only entries belonging to the same frequency range are marked at one time. The closer two homographic entries are in their frequency the greater the likelihood that they will both be marked and simultaneously available for the comparison process. (*d*) Nonsense words have longer RTs since the decision that the stimulus is nonsense requires exhaustive search, that is, marking down to the lowest frequencies, several comparisons and probably rechecking of the quantization outputs.

References

Riegel, K. F., & Riegel, R. M. Prediction of word-recognition thresholds on the basis of stimulus parameters. *Language and Speech*, 1961, **4**, 157–170.

Spreen, O., Borkowski, J. G., & Benton, A. L. Auditory word recognition as a function of meaningfulness, abstractness and phonetic structure. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 101–104.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.

Winnick, W. A., & Kressel, K. Tachistoscopic recognition thresholds, paired-associate learning, and free recall as a function of abstractness–concreteness and word frequency. *Journal of Experimental Psychology*, 1965, **70**, 163–168.

(Received February 13, 1970)

Appendix

Reaction Times (msec) of English Stimuli[a]

Nonhomographs, Concrete

Word Frequency 10

pond	744	tepid	908	filth	962	denim	1146
lung	767	moth	912	chef	968	foyer	1195
disk	851	merge	931	broom	1001	twig	1214
pest	855	ramp	931	shawl	1013	turf	1218
fern	908	sewer	958	shrub	1141	cove	1325

Word Frequency 100

cliff	707	grass	800	tray	860	gift	958		
lamp	756	clay	802	song	897	bacon	960		
barn	771	rice	817	movie	929	gang	970		
flesh	787	ocean	829	wagon	931	wine	1031		
bird	797	roof	854	cigar	940	wrist	1060		

Word Frequency 1000

chair	697	thing	732	lady	773	white	878
city	703	black	737	brown	815	money	892
girl	704	green	767	large	848	town	904
food	708	child	767	small	865	woman	926
dark	727	wife	770	door	877	word	958

Nonhomographs, Abstract

Word Frequency 10

risky	810	veto	947	cult	1017	shirk	1113
myth	856	glee	951	whim	1051	greed	1130
mirth	891	defy	967	smug	1059	lowly	1144
trite	897	elite	984	trait	1061	bias	1170
pact	940	farce	1004	zeal	1091	wary	1441

Word Frequency 100

loss	691	fury	802	grief	858	skill	923
pure	742	ugly	820	hint	860	solve	980
task	762	false	827	occur	865	lazy	989
legal	765	debt	850	cruel	870	fate	1037
lend	795	magic	856	event	904	quit	1171

Word Frequency 1000

also	679	soon	733	same	784	quite	831
happy	695	less	754	full	786	half	873
true	708	today	762	often	790	next	951
year	718	fact	764	idea	793	able	1029
alone	721	until	780	along	804	above	1072

Homographs, Concrete

Word Frequency 10

dice	800	crane	869	tart	972	stud	1060
fuse	814	rake	904	helm	973	crumb	1124
grill	819	stain	906	snuff	991	gorge	1241
		sling	940	mesh	1002		

Word Frequency 100

lace	758	sheet	805	pipe	832	organ	956
rock	762	chest	825	shed	859	blow	992
		belt	831	duck	948		

Word Frequency 1000

light	781	place	818	water	819	table	919

Homographs, Concrete/Abstract

			Word Frequency 10				
peck	849	raft	882	pawn	990		
flap	859	prone	887	pluck	1043		
			Word Frequency 100				
trace	716	dirt	747	block	826	alert	888
yard	744	charm	785	dumb	826	grave	1028
		nerve	818	star	838		
			Word Frequency 1000				
post	652	still	729	care	773	fire	832
late	688	party	739	miss	775	case	833
power	690	kind	765	head	809	left	850
feel	729	mind	770	mean	822	hard	941

[a] These are mean RTs of *correct* responses. Each mean was calculated by summing the logarithmic transforms of the RTs to a given word, dividing by the number of *S*s (39), and finding the antilog of this quotient. The divisor was constant since the RT of an incorrect response was replaced by the *S*'s mean RT for the class to which the stimulus belonged. This was done so that each word mean would be derived from the same group *S*s. Since it was the slower *S*s for whom the replacements generally had to be made, the effect of this adjustment was to raise the RTs. Thus, class means calculated from these word means will be consistently greater than those given in Table 1.