Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 4, 687–693

0278-7393/88/$00.75

Word Frequency and Generation Effects

John M. Gardiner
Memory & Cognition Research Group
City University, London, England

Vernon H. Gregg
Birkbeck College
University of London, England

James A. Hampton
Memory & Cognition Research Group
City University, London, England

Nairne, Pusen, and Widner (1985) described an experiment in which they found that, like nonwords, low-frequency words were no more likely to be recognized if they had been self-generated at study than if they had been read. These findings led Nairne et al. to conclude that generation effects depend on the extent to which words are associatively linked to other words in semantic memory, not on lexical representation per se. In a series of experiments designed to test the generality of Nairne et al.'s low-frequency finding, we found not only that generation effects occurred for low-frequency words but also that those effects were on the whole indistinguishable from effects obtained with high-frequency words. These findings indicate that word frequency does not define an essential boundary condition for generation effects and so provides no basis for preferring an associative hypothesis to a lexical one.

Generation effects may be defined, somewhat loosely, as superior memory performance with respect to test items that at study were self-generated rather than read. Such generation effects have received much attention and gained considerable theoretical importance since the publication of an article by Slamecka and Graf (1978), in which these effects were magisterially delineated and—on the not unreasonable grounds that no theory of memory seemed to offer a ready interpretation of them—left unexplained.

The search for an explanation of generation effects has naturally involved attempts to determine their boundary conditions, and a significant breakthrough in this respect was the discovery by McElroy and Slamecka (1982) that, at least in tests of item recognition and recall, no generation effects occurred with nonwords. This finding, which was soon confirmed and extended by other researchers (e.g., Nairne, Pusen, & Widner, 1985; Payne, Neely, & Burns, 1986), led McElroy and Slamecka to propose what later researchers called the *lexical activation hypothesis*. According to this hypothesis, generated items must be represented in the mental lexicon for a generation effect to occur and, compared with the read condition, the generate task enhances activation of the lexical representation; it is this enhanced activation that subsequently benefits performance.

Nairne et al. (1985) argued that representation in the mental lexicon was not a sufficient condition for generation effects to occur, and they proposed instead what we call the *associative linkage hypothesis*. According to this hypothesis, it is not an item's lexicality that is crucial but the extent to which it is linked associatively with other items in semantic memory. The reason that nonwords do not give rise to generation effects is, therefore, their lack of associative links and not the absence of lexical representation. In support of this alternative hypothesis, Nairne et al. reasoned that because low-frequency words have fewer associates than high-frequency words, the generate–read variable should interact with word frequency, and in fact they found no generation effect in the recognition of low-frequency words.

Drawing a sharp distinction between representation in lexical memory and representation in semantic memory, Gardiner and Rowley (1984) and Gardiner and Hampton (1985) also argued against the lexical activation hypothesis, but on the grounds that it specified the critical nature of generated items too narrowly. Essentially the claim was that any functionally integrated and familiar concept represented in semantic memory could give rise to generation effects. In support of this hypothesis, which we may call the *conceptual integration hypothesis*, Gardiner and Hampton showed that generation effects in recall occurred for meaningful but not meaningless letter bigrams (e.g., *E T* vs. *E C*), for integrated but not separate pairs of digits (e.g., *32* vs. *3, 2*), and for familiar but not unfamiliar noun compounds (e.g., *tomato ketchup* vs. *cheese ketchup*). According to this hypothesis, generation effects should depend neither on lexicality nor on associative linkage but simply on an integrated conceptual representation in semantic memory. Hence, contrary to the Nairne et al. (1985) result, generation effects should not depend critically on word frequency.

All three of these hypotheses attempt to fine tune the boundary conditions of generation effects by defining those conditions in terms of representational properties of the stimulus items. They therefore all share a common limitation: They do not take into account the relationship between study and test conditions, as encapsulated by the encoding specificity principle or by the rather similar principle of test appropriateness. This limitation is underscored by Nairne and

We thank James Nairne and David Payne for their helpful correspondence and permission to refer to their work in hand, and we thank Ruth Chandler and Caroline Tjoa for their help with materials preparation and data collection and analysis.

Correspondence concerning this article should be addressed to John M. Gardiner, Memory & Cognition Research Group, City University, Northampton Square, London, England, EC1V 0HB.

Widner's (1987) demonstration that, provided tests are constructed appropriately with respect to study conditions, generation effects do occur for nonwords (see also Johns & Swanson, 1988). Nairne and Widner (1987) demonstrated such effects by designing recognition tests that tapped memory for the specific operations carried out in a letter reversal task at study, rather than for the occurrence of the items in the study list. That generation effects occur for nonwords under certain study–test conditions clearly restricts the generality of all hypotheses that are stated without respect to those conditions.

Moreover, these developments, together with all other extant theories about generation effects, may now only be of historical interest—at least, if data and argument recently advanced by Slamecka and Katsaiti (1987) were to be unreservedly accepted.

In what could be regarded as a dramatic bid to keep generation effects unexplained, Slamecka and Katsaiti (1987) argued that these effects are merely an artifact of selective displaced rehearsal. That is, when generate and read items occur together in a mixed list, subjects tend to rehearse generated items at the expense of read items. In support of this claim, Slamecka and Katsaiti (see also Begg & Snider, 1987) showed that there were no generation effects in recall with a between-lists, as opposed to a mixed-lists, design and that when selective rehearsal of generate items was prevented by requiring subjects to rehearse only the current item, there were no generation effects within a mixed-lists design either. On the grounds that virtually all published studies of generation effects had used a mixed-lists design, Slamecka and Katsaiti argued that their data render all extant theories untenable.

It is possible that this claim may be exaggerated. One limitation of the selective rehearsal account is that it is not clear why subjects should adopt a selective rehearsal strategy and why their deployment of this strategy should depend so critically on the nature of the study materials—although Slamecka and Katsaiti (1987) did suggest that nonwords, or nonintegrated compounds such as *cheese ketchup*, may give rise to a current-item-only mode of study. Also, this account suffers from the same limitation as do the lexical activation, conceptual integration, and associative linkage hypotheses: It does not readily explain interactions between study and test conditions. Generation effects have been implicated in such interactions by a large number of studies (e.g., Gardiner, 1987, 1988; Glisky & Rabinowitz, 1985; Graf, 1980; Jacoby, 1983; Nairne & Widner, 1987; Rabinowitz & Craik, 1986; Roediger & Blaxton, 1987; Roediger & Weldon, 1987). And these studies converge on or are consistent with the conclusion that generation effects may comprise at least two encoding components, a data-driven or surface-processing component (see also Nairne, 1988), as well as a conceptually driven or semantic-processing component. It is hard to see how a selective rehearsal account, or for that matter any single factor account, might successfully be extended to encompass all these findings. And so perhaps other theoretical ideas about generation effects are not yet merely of historical interest.

There now seems little doubt that generation effects are quite complex. Hirshman and Bjork, for example, proposed another two-factor account of generation effects, according to which generating a response item leads to deeper encoding of both the response itself and the relation between the response and the stimulus with which it is paired. They showed that in line with this account, generation effects in free recall can be dissociated from generation effects in cued recall. And in contrast with Slamecka and Katsaiti (1987), Hirshman and Bjork obtained generation effects by using between-subjects designs, although their results suggested that mixed-lists designs inflate generation effects because of differential allocation of attention to generate and read items. Begg and Snider (1987) also hypothesized that in mixed-lists designs the generate task induces a more cursory encoding of read items. And they too compared mixed with unmixed lists and found no generation effects with unmixed lists in recognition tests. But then that result conflicts with results reported by Slamecka and Graf (1978), who obtained generation effects with a between-subjects design and recognition tests.

In the face of such a perplexing pattern of results, it is tempting to conclude that generation effects have gone from being unexplained to being inexplicable. However, no doubt these matters will be clarified in due course. In the meantime, our purpose in the present article is relatively modest and straightforward: It is to provide further evidence relevant to an evaluation of the lexical activation, associative linkage, and conceptual integration hypotheses. Although clearly limited in scope, these hypotheses continue to be of interest because the relation between generation effects and the semantic properties of items remains an open question. More particularly, our major aim was to test the generality of the Nairne et al. (1985) finding of no effect for low-frequency words. This finding is important because it is the only evidence that supports the associative linkage hypothesis and refutes both lexical activation and conceptual integration hypotheses.

Experiment 1

In the experiment of interest, Nairne et al. (1985, Experiment 3) presented their subjects with a 48-item study list consisting of 12 nonwords and 12 words each of low, medium, and high frequency. Half the items were generated, half read; for the generate task, the first two letters of the item were shown in reverse order and underlined, so that the task entailed simply reordering the letters. In a yes–no recognition memory test, Nairne et al. found no generation effects for the nonwords or low-frequency words, but there were generation effects for the medium- and high-frequency words. The purpose of Experiment 1 was merely to replicate the crucial finding, the absence of a generation effect with low-frequency words.

Method

Subjects. The subjects were 24 undergraduate students at Birkbeck College, London, who were tested, during a laboratory course, in two equal groups to which subjects were allocated arbitrarily.

Design and materials. The subjects studied a list of 24 low-frequency words, half of which were generated and half read. Generate words for one group of 12 subjects were read words for the other.

The 24 low-frequency words consisted of the 12 words used by Nairne et al. (1985, Table 1) plus a further 12 words that were matched with them for frequency, imagery values, and word length. Their average frequency, from the Kucera and Francis (1967) norms, was 2.75 for Nairne et al.'s words and 2.17 for ours; their average imagery value, from the Paivio, Yuille, and Madigan (1968) norms, was 3.99 for both sets of words. About 80 min after presentation of the study list, the subjects were given a yes–no recognition test in which the 24 target words were intermixed with 24 lure words of a similar nature.

Procedure. Study-list words were presented in uppercase letters from an overhead projector at the rate of 5 s per word. Two versions of a single random order of study-list items were used; generate items in one version were read items in the other. For the generate task, the first two letters of each word were inversely ordered and underlined. Subjects were instructed to write down all the words, as words, in a column, using a cardboard mask such that only the word currently being copied was visible. The subjects were informed at study that there would subsequently be a memory test, but they were not told what sort of test it would be. About 80 min after the study-list presentation (a retention interval in which subjects were occupied in attempts to master some statistical techniques), recognition test sheets were distributed. The 48 test words, 24 targets and 24 lures, were typed in four columns of 12 in a constant random order. The subjects were instructed to work carefully down each column in turn, drawing a circle around any word they recognized from the study list.

Results and Discussion

At study there was a 3% failure rate in the copying task, 10 cases in the generate condition and 8 in the read condition; these items were excluded from data analysis. In the recognition test, a number of subjects attained the maximum possible number of hits—11 for read items and 14 for generate items. For this reason, the main results, which are summarized in the first two rows of Table 1, are shown separately both for uncorrected hit rates and for hit rates corrected by subtracting false positive rates. For the same reason, separate statistical analyses were carried out on uncorrected and corrected scores, but because the statistical conclusions did not differ, only the results for corrected scores are reported. These results show that generate items were significantly more likely to be recognized than read items, $t(23) = 2.97$, $p < .01$, $SE = .04$. An analysis of variance (ANOVA) that included word source as a variable was also carried out; that is, the 12 low-frequency words used by Nairne et al. (1985) were compared with the 12 matched words that we added. There was a significant generation effect, $F(1, 23) = 5.30$, $p < .05$, $MS_e = .03$, in this analysis too, but neither word source nor the interaction between word source and task approached significance, with $F < 1$ in each case. Somewhat to our consternation, therefore, our results conflict with those of Nairne et al. by demonstrating a generation effect with low-frequency words—including the words used in that study. This outcome offers no support to the associative linkage hypothesis. Instead, it is completely consistent with McElroy and Slamecka's (1982) lexical activation hypothesis and with the conceptual integration hypothesis advanced by Gardiner and Rowley (1984) and Gardiner and Hampton (1985).

Experiment 2

Our purpose in Experiment 2 was to replicate the finding of a generation effect with low-frequency words and, in order to provide a further and stronger test of the associative linkage hypothesis, to manipulate word frequency. In addition, Experiment 2, and also Experiments 3 and 4, involved further tests of generality with respect to materials: In each experiment different samples of words were used.

Table 1
Recognition Scores for Experiments 1–3

Experiment	Read	Generate	Difference
Experiment 1			
Low frequency			
Uncorrected	.79	.90	
Corrected	.71	.82	.11
Experiment 2			
Low frequency			
Uncorrected	.79	.88	
Corrected	.76	.85	.09
High frequency			
Uncorrected	.56	.81	
Corrected	.51	.76	.25
Experiment 3			
Low frequency			
Uncorrected	.66	.77	
Corrected	.53	.64	.11
High frequency			
Uncorrected	.54	.67	
Corrected	.35	.48	.13

Note. Uncorrected scores are raw hit rates; corrected scores are hit rates minus false positive rates.

Method

Subjects. The subjects were 24 undergraduate student volunteers at the City University, London, who were paid for their participation. They were tested in four groups with 6 subjects allocated arbitrarily to each group.

Design and materials. A total of 96 words, 48 words each of low and high frequency, were taken from the Medical Research Council (MRC) Psycholinguistic Database (described by Coltheart, 1981). The low- and high-frequency words selected had mean frequency counts of 2 and 160 in Kucera and Francis (1967). Both low- and high-frequency words had the same high average imagery values of 609, as calculated in the data base by merging three sets of norms to form a scale ranging from 100 to 700. The 96 words were randomly ordered in a single constant order for the recognition test. The study lists consisted of 24 low-frequency and 24 high-frequency words. There were two different study lists, such that list words for half the subjects were lure words in the recognition test for the other half. There were equal numbers of 1-, 2-, and 3-syllable words at each frequency level, and words of each syllabic length were assigned to each study list in equal proportions. There were two versions of each study list, such that generate words for half the subjects were read words for the other half. There was a 20-min retention interval, because it was thought that, relative to Experiment 1, the doubling of list length would offset the effects of reducing the retention interval.

Procedure. The procedure was essentially the same as in Experiment 1. Presentation rate, however, was slightly slower at 6 s per word. The yes–no recognition test sheets consisted of four columns of 24 words. During the 20-min retention interval, the subjects were engaged in a semantic memory experiment about disjunctive concepts.

Results and Discussion

In the copying task, there was only one word that was copied incorrectly. The main results are summarized in Table

1, both for uncorrected recognition scores and for scores corrected by subtracting false positive rates from hit rates. In this experiment, fewer subjects attained the maximum possible number of hits: for low-frequency words, 3 for read and 5 for generate items; for high-frequency words, 1 for read and 4 for generate items. Separate statistical analyses again yielded the same conclusions for uncorrected and corrected scores, so only the results of the ANOVA for corrected scores are reported. These results show that both the main effects of word frequency (the standard word frequency effect in recognition memory; see Gregg, 1976), $F(1, 20) = 32.25$, $p < .001$, $MS_e = .02$, and of task, $F(1, 20) = 33.26$, $p < .001$, $MS_e = .02$, were significant. There was also a significant interaction between them, $F(1, 20) = 20.39$, $p < .001$, $MS_e = .01$. A separate analysis of generate and read conditions for low-frequency words only, however, revealed a significant generation effect, $F(1, 20) = 8.74$, $p < .01$, $MS_e = .01$. Thus both word frequency and generation effects were replicated in this experiment, and although the size of the generation effect was reduced with low-frequency words, it was nonetheless still reliably present.

These findings, although they again fail to replicate those of Nairne et al. (1985), are consistent with the associative linkage hypothesis, inasmuch as they show the predicted interaction between word frequency and generation effects. Nevertheless, it is possible that this interaction may have come about for reasons that have little or nothing to do with this hypothesis. For example, despite the use of much longer retention intervals than used by Nairne et al. (only 1 min), the level of recognition performance was still quite high in Experiments 1 and 2. This suggests that the interaction might reflect ceiling effects. Experiment 3 was designed therefore to provide evidence on the relationship between word frequency and generation effects when performance was appreciably lower.

Experiment 3

Experiment 3 was intended primarily to provide a straightforward replication of Experiment 2 with one major methodological change: The retention interval was increased to 24 hr.

Method

Subjects. The subjects were 16 graduate students and technical or secretarial staff at Birkbeck College, London, or Sussex University. They were paid for their participation in the experiment and tested individually or in groups of 2 or 3.

Design and procedure. The design and procedure were essentially the same as in Experiment 2, with two exceptions: The retention interval was about 24 hr, and there was another sampling of 96 words, 48 words each of low and high frequency, taken from the MRC Psycholinguistic Database (Coltheart, 1981). In this sample the mean respective frequencies, according to Kucera and Francis (1967), were 1.36 and 1.30; the imagery values were 617 in each case. Words at each frequency level were again balanced with respect to their syllabic length.

Results and Discussion

Subjects' performance in the copying task was error free. Table 1 summarizes the main results of the experiment in the same manner as before. Performance was clearly lower than in Experiment 2, but there were still some subjects who attained the maximum possible hit rate: 5 for low-frequency and 2 for high-frequency generate items. Statistical analyses are again reported only for corrected scores, because separate analysis of uncorrected and corrected scores yielded similar outcomes. Both main effects were significant: $F(1, 15) = 21.89$, $p < .001$, $MS_e = .02$, for word frequency, and $F(1, 15) = 13.37$, $p < .01$, $MS_e = .02$, for task; the interaction was not significant ($F < 1$), thus providing no support for the associative linkage hypothesis. These results suggest that the interaction observed in Experiment 2 might not be relevant to the hypothesis. Why might this be?

From inspection of the data shown in Table 1, apparently it is the generation effect for the high-frequency words in Experiment 2 that is exceptional and gives rise to the interaction in that experiment. Were the interaction in Experiment 2 to be attributable to ceiling effects, one would expect its disappearance in Experiment 3 to be associated with a larger generation effect for the low-frequency words in Experiment 3, but clearly this did not happen. Also, there was no support for a ceiling effect interpretation in a post hoc median split of data from Experiment 2 between high- and low-scoring subjects. This pattern of results and the failure to replicate the interaction also provide little support for an interpretation of the interaction in terms of associative linkage, unless it could be shown that other properties of the words selected for the two experiments differ in a manner predicted by the hypothesis and so can be used to explain the discrepancy between them. Word frequency is of course only one index of associative linkage, and an imperfect one at that, as Nairne et al. (1985) acknowledged. Meaningfulness measures would be a more direct index. Also, Nairne and Widner (1987) suggested that perhaps in recognition, though not in recall, familiarity might be more important than associative linkage. In any event, Table 2 summarizes meaningfulness and familiarity values for all those words used in Experiments 2 and 3 for which norms were available in the MRC Psycholinguistic Database (Coltheart, 1981). Values in these norms range from 100 to 700, a range that is equivalent to that of 1 to 7 in the original norms from which the scale was derived. It is clear from Table 2 that within each frequency band the words used in Experiments 2 and 3 had virtually identical meaningfulness and familiarity values. The finding of an interaction between word frequency and generation effects in Experiment 2 but not in Experiment 3 cannot therefore be related to differences in these other characteristics of the words. The interaction in

Table 2
Meaningfulness and Familiarity Ratings for Words Used in Experiments 2 and 3

	Low frequency				High frequency			
Exp.	Meaningfulness	No.	Familiarity	No.	Meaningfulness	No.	Familiarity	No.
2	448	31	487	46	507	38	584	48
3	440	36	485	47	511	42	587	47

Note. Exp. = experiment. No. = number of words used for which ratings were available in the norms (Coltheart, 1981); in these norms, values range from 100 to 700.

Experiment 2 may therefore be a spurious result—possibly a rehearsal strategy artifact. Alternatively, there may be a genuine second-order interaction between task, word frequency, and retention interval, although of course to confirm this would require further research.

Experiment 4

Our final experiment was originally conceived as part of a planned set of studies in which different dimensions of stimulus materials were to be tested for their possible influence on generation effects. We report it here because, although not a direct test of the different hypotheses previously described, it speaks directly to the central empirical question of whether low-frequency words support a generation effect equivalent to that for high-frequency words. The experiment is also relevant to evaluating a possible criticism of our attempts to replicate the findings of Nairne et al. (1985), concerning the use of recognition testing throughout. As mentioned in the introduction, there is evidence for generation effects that may reflect data-driven rather than conceptually driven processing. The generation rule used in the present experiments involved the extremely simple surface-type task of reversing two letters. This task might have failed to engage any deeper semantic processing of the generated item. Some theorists assume that recognition memory involves a data-driven component (e.g., Jacoby & Dallas, 1981; Mandler, 1980), so it is possible that the generation effects obtained in these recognition tests depended on data-driven processing, or the retention of additional surface or perceptual information. And if that is so, then the effects are not relevant to hypotheses about the semantic properties of the items. There is evidence that free recall involves little or no data-driven processing (see, e.g., Roediger & Weldon, 1987), so an appropriate way to check this possibility would be to see whether the letter reversal task produces generation effects in recall.

Method

Subjects. The subjects were 40 unpaid volunteers, some of whom were undergraduate students at the City University, London, and others of whom were drawn from the public. They were allocated arbitrarily to one of two groups of 20 subjects and were tested individually.

Design and materials. One group of 20 subjects learned a list of 24 low-frequency words; the other group learned a list of 24 high-frequency words. In each case 12 of the words were generated and 12 read, and words designated as generate for half the subjects within each group were read for the other half. The low-frequency list averaged 3.9 by the Thorndike–Lorge (1944) word count, and all high-frequency words were designated A or AA. The average imagery values were 4.4 and 4.5, and average meaningfulness values were 5.3 and 5.6 for low- and high-frequency words, respectively (values taken from Paivio et al., 1968). Word length was equated in terms of the average number of letters, which was 6.4 for low- and 6.3 for high-frequency words. There were three free-recall learning study and test trials.

Procedure. The generate task was the same as in the preceding experiments. The study-list words were presented one at a time through a deck of cards, which were shuffled anew for each subject and for each trial. The subjects were required to say each word aloud as it was presented, rather than copy words down, and the presentation rate was 5 s per word. At the end of each study-list presentation, subjects were given a card showing a 3-digit number that they had to repeat aloud and count aloud from, backward by threes. This task lasted for 30 s, and its purpose was to reduce the level of recall of recency items. Free recall was written and typically took about 2 min.

Results and Discussion

The principal results of this experiment are summarized in Figure 1. The results of an ANOVA carried out on individual subjects' recall scores confirmed the most important findings that are apparent from the figure, which are the presence of generation effects, $F(1, 38) = 10.18$, $p < .01$, $MS_e = .03$, and the absence of an interaction between word frequency and generation effects ($F < 1$). The word-frequency effect itself (i.e., the superior recall of high-frequency lists) was not quite significant, $F(1, 38) = 2.37$, $p = .13$, $MS_e = .10$, which may partly reflect the use of a relatively insensitive between-subjects comparison. The effect of trials was highly significant, $F(2, 76) = 179.30$, $p < .001$, $MS_e = .01$. Neither the Word Frequency × Trials interaction nor the three-way interaction was significant ($F < 1$ in each case). The apparent attenuation of the generation effect over trials did not quite result in a significant interaction, $F(2, 76) = 2.19$, $p < .12$, $MS_e = .02$. A separate ANOVA carried out on data for low-frequency words confirmed that there was a significant generation effect, $F(1, 19) = 6.80$, $p < .025$, $MS_e = .02$; the effect of trials, $F(2, 38) = 82.01$, $p < .001$, $MS_e = .01$, was also significant; the interaction was not ($F < 1$).

The finding of similar generation effects for high- and low-frequency words in free-recall learning is consistent with the idea that the letter reversal task does engage deeper conceptual processing, and so argues that the similar effects observed in the recognition experiments did not result from the retention of perceptual information.

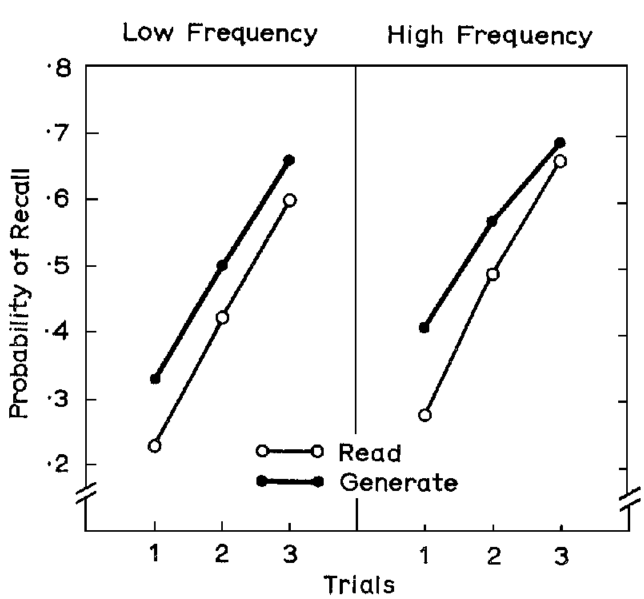

Figure 1. Mean recall probabilities for Experiment 4.

General Discussion

The main results from this series of experiments are that contrary to the result reported by Nairne et al. (1985), generation effects occurred in recognition memory for low-frequency words. Our results are therefore consistent with the lexical activation and the conceptual integration hypotheses but provide no support for the associative linkage hypothesis. Although the generation effect in Experiment 2 was found to be greater for the high-frequency words, as predicted by the associative linkage hypothesis, this interaction was not replicated in Experiment 3, and on closer analysis it was shown not to be readily interpretable in terms of associative linkage. The absence of any general relation between word frequency and generation effects was further confirmed in Experiment 4, which provided convergent evidence that the letter reversal task did engage deeper conceptual processing. This result renders it unlikely that the generation effects in recognition memory were not relevant to the hypotheses under evaluation because they reflected only the retention of perceptual or data-driven information.

The discrepancy between the Nairne et al. (1985) result and our own findings leads obviously to the question of why the generality of their null effect should be so constrained. The results of two other sets of studies bear directly on this question. Payne, Grosofsky, and Waring (1987) also found generation effects in recognition memory for low-frequency words and, moreover, in a direct replication of the Nairne et al. experiment—that is, with a study list composed of nonwords and words of low, medium, and high frequency and a retention interval of 1 min. Payne et al. (1987) found too that generation effects occurred for low-frequency words when nonwords were omitted from the study list. Thus the results of the Payne et al. (1987) studies replicate our own results and suggest that neither list composition (the presence of nonwords in the list might have reduced the extent to which low-frequency words were recognized and encoded as words) nor retention interval was a crucial factor.

Fittingly enough, it is Nairne and Widner (1988) themselves who, in the second set of studies mentioned, identified the crucial factor. Nairne and Widner (1988) hypothesized that their previously obtained null effect for low-frequency words might reflect familiarity, rather than frequency, so they presented a list of low-frequency words that were either of high or of low familiarity, together with nonwords. In addition, they had subjects indicate whether each presented item was a word or a nonword, thereby providing an explicit check on lexical representation and encoding. Their results were clear-cut: High-familiarity words gave rise to generation effects, low-familiarity words and nonwords did not, even though subjects correctly recognized the low-familiarity words as words. So Nairne and Widner (1988) replicated both the original Nairne et al. (1985) low-frequency result and the results obtained both by us in this article and by Payne et al. (1987). In so doing they identified familiarity as the critical variable. Their results constitute further evidence against the lexical activation hypothesis but are consistent with the other two.

The conclusion from all of this research is relatively straightforward. Word frequency does not itself define an essential boundary condition for generation effects, so it provides no grounds either for rejecting the lexical activation hypothesis or for supporting the associative linkage hypothesis. The evidence overall does not support the lexical activation hypothesis, however, but instead favors a broader semantic memory view of what the essential properties of the stimulus materials are. We believe that these properties are captured better by the conceptual integration hypothesis than by the associative linkage hypothesis. But at the limit, if it were assumed that only a certain minimal number of associative links is critical, rather than the overall number, then it would be difficult, if not impossible, to distinguish empirically between the two hypotheses.

Finally, as pointed out in the introduction, all three hypotheses are limited in scope because generation effects appear to entail more than one encoding factor, and because the effects depend on test conditions. Of particular importance in defining the scope of these hypotheses is the further elucidation of those conditions under which generation effects occur for nonwords. In this connection, Johns and Swanson (1988) showed that small but reliable generation effects occur in recognition memory for nonwords, provided that the nonwords in the generate condition are presented visually during the generate attempt. Johns and Swanson argued that previous failures to observe generation effects with nonwords reflected a confounding with familiarity factors, because subjects not only lacked previous visual experience with the nonwords but also did not get to see them during presentation. Interestingly, this account therefore also emphasizes familiarity, but familiarity of surface features rather than semantic features. Perhaps for nonwords, familiarity with respect to surface features plays a role parallel to that of semantic familiarity for words.

References

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 553–563.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497–508.

Gardiner, J. M. (1987, July). *Enhanced priming with generate rather than read study conditions.* Paper presented at the joint meeting of the Experimental Psychology Society and the Experimental Division of the Canadian Psychological Association, Oxford, England.

Gardiner, J. M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 495–501.

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 732–741.

Gardiner, J. M., & Rowley, J. M. C. (1984). A generation effect with numbers rather than words. *Memory & Cognition, 12*, 443–445.

Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 193–205.

Graf, P. (1980). Two consequences of generating: Increased inter- and intraword organization of sentences. *Journal of Verbal Learn-*

ing and Verbal Behavior, 19, 316–327.

Gregg, V. H. (1976). Word frequency in recognition and recall. In J. Brown (Ed.), *Recognition and recall* (pp. 183–216). London: Wiley.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 180–190.

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

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

McElroy, L. A., & Slamecka, N. J. (1982). Memorial consequences of generating nonwords: Implications for theories of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21,* 249–259.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 248–255.

Nairne, J. S., Pusen, C., & Widner, R. L., Jr. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition, 13,* 183–191.

Nairne, J. S., & Widner, R. L., Jr. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 164–171.

Nairne, J. S., & Widner, R. L., Jr. (1988). Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 694–699.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs, 76*(1, Pt. 2).

Payne, D. G., Grosofsky, A., & Waring, S. M. (1987). *The word–nonword difference in the generation effect: Lexicality, frequency, or familiarity.* Unpublished manuscript.

Payne, D. G., Neely, J. H., & Burns, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition, 14,* 246–252.

Rabinowitz, J. C., & Craik, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language, 25,* 226–237.

Roediger, H. L., III., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III., & Weldon, M. S. (1987). Reversing the picture superiority effect. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related processes: Theories, individual differences, and applications* (pp. 151–174). New York: Springer-Verlag.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26,* 589–607.

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

Received October 15, 1986
Revision received November 24, 1987
Accepted December 2, 1987 ■