Bulletin of the Psychonomic Society
1974, Vol. 4 (2A), 93-94

Word frequency in problem solving*

CARL P. DUNCAN
Northwestern University, Evanston, Illinois 60201

In a modification of an earlier study, Ss were presented with problems consisting of five words varying in printed language frequency. In different groups, Ss tried to guess the word at a particular frequency level in each problem. The words used within a problem were structurally more similar to each other than had been the case in the previous study. The result was that, in the present experiment, the word-frequency dimension emerged even more clearly as a habit strength variable in problem solving.

The response hierarchy approach to problem solving (Duncan, 1967) has generated a number of studies in which S must guess the correct response from among several responses arranged in a hierarchy of habit strength. Duncan (1967) used words varying in strength of association to a stimulus word, Dominowski (1965) had bigrams of different frequencies of occurrence in the printed language. In all of these studies the results were essentially the same. First, the habit hierarchy exerted a linear effect on correctness of response guessing such that stronger habit strength items (e.g., higher frequency words) were easier to guess than weaker items. Second, there was a quadratic effect in that both stronger and weaker responses were easier to guess than responses of intermediate strength. Thus, among sets of five words varying in frequency, used by Jurca and Duncan (1969), the most frequent word was easiest to guess, the fifth and fourth most frequent were hardest to guess. The quadratic effect was more powerful than the linear effect.

The present study is a modification of the Jurca and Duncan experiment. In that earlier study, S was presented with a number of problems. Each problem consisted of five five-letter words. The five words varied in Thorndike-Lorge (1944) frequency from less than 1 per million to 50 or more per million. Each of the five words within a problem began with a different letter; final letters varied haphazardly. In one condition, Ss were told in advance that word frequency was the critical dimension. These Ss found the task easier than uninformed groups, and they also showed a more clear-cut quadratic effect. It is possible that characteristics of the words other than their frequency level contributed to these results. If the differing initial letters, and usually differing final letters, served as cues, then guessing behavior was determined in part by factors other than word frequency. The present study is an attempt to check on this by using problems in which the words are structurally more similar to each other, so the effect of the frequency dimension should be clearer.

*Thanks are due Donald Lehr, John Perrachione, and Constance Duncan.

METHOD

Materials

Each problem consisted of five five-letter words. The words represented five levels of printed language frequency, varying from very low (1 or less per million) to very high (50-100 or more per million). All of the five words in a problem began with the same consonant and ended with the same consonant (initial and final consonants were different). There were 20 such problems, each using a different initial consonant for the words.

The five words for a problem were typed from left to right on a 4 x 6 in. card. Five different left-to-right orders were used for each problem, none of them corresponding to the frequency dimension.

Procedure

The S was told that he would be asked to guess the (1st, 2nd, 3rd, 4th, or 5th) most frequent word from among a set of five words. E explained the concept of printed language frequency of words until he was sure S understood it, and that guessing was to be based on the frequency dimension alone. There were five groups of Ss, one for each of the five levels of frequency. The Ss within a group tried to guess the level of frequency assigned to them on all 20 problems. The E told S whether he was right or wrong on each guess (problem), but did not tell S the correct word on wrong guesses. There were 60 Ss in each of the five groups.

RESULTS AND DISCUSSION

The mean number of correct guesses, in the order of lowest to highest word frequency levels, was 10.3, 8.3, 11.4, 11.6, 15.2. Thus, as might be expected, the highest frequency words were easiest to guess (15.2 out of 20 problems), while the most difficult (8.3) was the 4th, next to the lowest, frequency. Analysis of these means for algebraic trend showed a very significant linear trend: $F(1,295) = 176.1$, $p < .001$. The quadratic trend was also significant: $F(1,295) = 44.5$, $p < .001$. In the Jurca and Duncan (1969) study, the F value for quadratic trend was considerably larger than the value for linear trend. Thus, the more highly structurally similar words used in the present study produced a reversal in that the most powerful effect was that higher frequency words were easier to guess than lower frequency words (linear effect).

In the Jurca and Duncan study, there were 15 problems. The mean number of correct guesses in that study was 10.3, 6.0, 6.5, 8.6, and 11.4, from lowest to highest frequency level. For the first 15 problems of the present study the corresponding means were 7.7, 6.2, 8.6, 8.7, and 11.4. At the highest frequency level, the means of the two studies are identical. In the Jurca and Duncan experiment, the lowest frequency level (mean = 10.3) was easier to guess than any level except the highest (thus a strong quadratic trend). In the present data, the "upturn" at the lowest level is not so sharp, and in fact that level is the most difficult to guess except for Level 4.

For each of the five groups, error gradients were plotted showing the number of errors made to each of the four wrong words over the 20 problems. These will not be presented here, but they are easily described. Without exception, more errors were made to words closer in frequency to the correct word than to words farther away on the frequency dimension. For example, Ss trying to guess the middle frequency word (Level 3) made more guesses (errors) to words at Levels 2 and 4 than at Levels 1 and 5. Error gradients of this kind are shown in Duncan's (1967) study of guessing words associated in differing strengths to a stimulus word.

In the Jurca and Duncan (1969) experiment, Ss were trying to guess words at a particular frequency level. However, the five words in a problem were selected so that they all began with different letters, and final letters varied haphazardly. Thus, it was possible for S to base his guess, at least in part, on letter cues rather than on word frequency. Such distractor cues might be either facilitating or interfering in S's attempts to maximize his score, but in any case such cues act as noise mixed in with the frequency dimension. In the present study, an effort was made to reduce such distractor cues by having the words in a problem begin with the same letter and end with the same letter. The main difference in results between the two studies is that the linear effect of word frequency on guessing (higher frequency words were easier to guess than lower frequency words) was the strongest effect in this study, whereas Jurca and Duncan found the quadratic effect stronger (terminal frequency levels easier to guess than intermediate levels). The "end effect" of terminal-level words is probably not completely avoidable when all words are present to S. But such error factors can be minimized, as the present study shows. When this is done, word frequency as a measure of habit strength emerges even more clearly as a powerful variable in problem-solving tasks.

REFERENCES

Dominowski, R. L. Problem difficulty as a function of relative frequency of correct responses. Psychonomic Science, 1965, 3, 417-418.

Duncan, C. P. Response hierarchies in problem solving. In C. P. Duncan, (Ed.), *Thinking: Current experimental studies.* Philadelphia: Lippincott, 1967. Pp. 18-41.

Jurca, N. H. & Duncan, C. P. Problem solving within a word-frequency hierarchy. Journal of Verbal Learning & Verbal Behavior, 1969, 8, 229-233.

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

(Received for publication May 13, 1974.)