

Semantic relatedness and clustering in free recall

MAY F. D'AMATO

Brooklyn College, City University of New York
Brooklyn, New York 11210

Nine replications investigated clustering in the immediate free recall of semantically related and semantically unrelated associates of low associative strength according to free association norms. The results showed more clustering for semantic than for nonsemantic associates of the same frequency and associative strength and revealed that this effect was not due to the increased recall of semantic associates.

It is now well established that categorically related free associates tend to cluster in free recall (e.g., Marshall, 1967). Thus, if subjects are presented with a list made up of randomly arranged members of pairs such as *man-woman*, *table-chair*, etc., they show a tendency to recall members of the pairs together. Since the pair members belong to a common category and are also associatively related it is not clear which of these two factors is responsible for the observed clustering. Associative explanations have often been favored (Bousfield, Puff, & Cowan, 1965). However, when Marshall (1967) matched lists of categorized pairs (e.g., *wool-silk*) and noncategorized pairs (e.g., *whistle-train*) for associative relatedness, he found that the categorized pairs clustered more than the noncategorized pairs at low to moderate associative strengths. Thus, an associative interpretation of clustering was shown to be incomplete.

Marshall did not control for frequency. The clustering differences he observed may therefore have been due to possible frequency differences rather than to differences in the extent of relatedness in meaning of list pairs. The present study seeks to determine whether associates related in meaning cluster more than associates not related in meaning when frequency is controlled.

METHOD

Materials

Two sets of lists of eight pairs of associated words were used. Each set consisted of two lists containing the same free association stimulus words. In one of the lists, the remaining eight words consisted of semantically related free associates to each stimulus word (e.g., *lamb-goat*), while in the other list these words were nonsemantic associates (e.g., *lamb-stew*). Pairs were selected such that the associative strength and frequency were approximately equal, insofar as possible, for semantic and nonsemantic associates to each stimulus word. This procedure differs from the one employed by Marshall which did not use the same free association stimulus words in categorized and noncategorized lists. The stimulus words for the first set of lists were *baby*, *yellow*, *box*, *train*, *dumb*, *high*, *lamb*, and *soap*. The

This paper is sponsored by E. G. Heinemann who takes full editorial responsibility for it.

semantic and nonsemantic associates, respectively, to each of these words were *infant*, *crib*; *brown*, *sun*; *crate*, *cardboard*; *engine*, *fast*; *ignorant*, *bell*; *tall*, *mountain*; *goat*, *stew*; and *suds*, *dish*. The stimulus words for the second set of lists were *silk*, *tobacco*, *shoulder*, *square*, *green*, *insect*, *scared*, and *cold*. The semantic and nonsemantic associates, respectively, to each of these words were *satın*, *scarf*; *cigar*, *smell*; *elbow*, *harness*; *shape*, *dance*; *purple*, *apple*; *spider*, *pest*; *brave*, *stiff*; and *sick*, *snow*. With but few exceptions the difference in associative probability between semantic and corresponding nonsemantic associate was no more than .01 based on the Bilodeau and Howell (Note 1) and the Palermo and Jenkins (1964) norms, while the difference in frequency (Kucera & Francis, 1967) did not exceed 8. The mean and range of associative probabilities for the first semantic and nonsemantic lists were .041, .02-.08 and .045, .02-.07, respectively. These values for the second semantic and nonsemantic lists were .038, .01-.05 and .036, .01-.04. Thus, the lists were well matched for associative probability as were semantic and corresponding nonsemantic pairs insofar as possible. For the first set the mean frequency for the semantic list was 40.1 with a range from 2-176. The mean frequency and range for the nonsemantic list were 34.0 and 5-112, respectively. For the second set the mean frequency and range for semantic and nonsemantic list, respectively, were 25.0, 2-85 and 28.9, 4-90. In selecting items care was taken to assure that direct forward associations existed only between the members of each pair and that only members of semantic pairs were related in meaning. The words in the lists were arranged in an order such that the stimuli always preceded their associates and that no associated words occurred together. The order of stimuli and their associates was identical for both semantic and added unrelated word (*proof*) at the end, while the second set had two such words (*idea* and *wall*).

Subjects

A total of 240 students from courses in introductory psychology, experimental psychology, and statistics served as subjects. They were tested in intact classes of from 8-38 individuals.

Procedure

In all, nine different classes were tested, each representing one replication of the study. The only difference among replications was in the set of lists used or in the order of the items within lists. Each list was dittoed and appeared on a single sheet of unlined paper which was covered by a card when handed out to each subject. Semantic and nonsemantic lists were alternated. The subjects exposed each word for 2 sec, in time to the experimenter's signal, by moving the card down the list aligning it with successive marks to assure proper exposure of each word in turn. Three practice items (nonsense syllables) were exposed

before the list itself. Prior to presentation of the list the subjects were given standard free recall instructions. No information was provided as to the nature of the relationship involved in each list. Immediately after all items had been exposed, the list was placed out of view and the subject was given a 2.5-min free-recall period.

RESULTS AND DISCUSSION

For each subject, the number of pairs of associated words recalled in forward clusters was determined. Forward clusters only were analyzed as matching for associative strength had been done only on the basis of the associative probability of forward direct associations since complete norms for the backward associations were not available. In each replication both the mean and median number of clusters for semantic associates were larger than for nonsemantic associates. The means for semantic and nonsemantic associates, respectively, for each of the nine replications were 1.33 and .54, 1.54 and 1.00, 1.75 and .25, 2.67 and .87, 1.08 and .92, 1.31 and .43, 1.08 and .71, 1.72 and .35, and 1.67 and 1.11. The medians for semantic and nonsemantic associates, respectively, were 1.30 and .58, 1.12 and .88, 1.17 and .33, 2.40 and .60, .83 and .67, 1.50 and .40, .88 and .64, 1.80 and .20, and 1.13 and 1.00. Since the data suggested that the underlying distributions were markedly nonnormal, a median test was performed on the clustering data of each replication. The resulting values of chi square, corrected for continuity, ranged from 0-8.53, yielding a combined $\chi^2(9) = 27.39$, $p < .01$.

The number of items recalled showed a small but consistent superiority for lists of semantic associates. Thus, it might be argued that the greater number of semantic clusters reflected the somewhat heightened recall of these types of items. Therefore, a second clustering measure was used, one which was independent of the amount recalled. This measure, the ratio of the number of items recalled in clusters to the total number of items recalled, revealed results similar to those involving merely the number of clusters. That is, without exception the means were greater for semantic associates. This was true for the medians as well with but one exception in which case the values were identical for semantic and nonsemantic associates. Median tests provided corrected chi-square values ranging from 0-8.53 for a combined $\chi^2(9) = 18.40$, $p < .05$. These results reveal that superior clustering existed for semantic associates independent of the number of items recalled.

Concerning recall, it has been mentioned that a small but consistent superiority existed for semantic associates. For each replication, the mean number recalled was greater for semantic associates ranging from 10.33-12.87 for an overall mean of 11.34. The range in mean number recalled for nonsemantic associates was

from 9.00-10.77 with an overall mean of 10.28. A t test performed on the recall data of each replication yielded t values that ranged from .44-3.40. The corresponding probabilities, when combined by the method suggested by Winer (1962, pp. 43-44), resulted in an overall $p < .005$. Thus, in the recall of weakly associated word pairs a small but significant superiority exists for lists of semantically related pairs.

In revealing that semantically related weak free associates cluster more than nonsemantic associates the present results show that Marshall's (1967) finding is a replicable phenomenon when frequency is controlled. In addition to matching the frequency of the semantic and nonsemantic free associates to each of the free association stimuli this study employed the same free association stimuli in both semantic and nonsemantic lists. Thus, the increased clustering of semantic associates cannot be attributed to frequency differences in the responses or availability differences in the stimuli.

While the results of this study rule out associative relatedness as the sole determiner of clustering, the experiment does not permit a conclusion as to how semantic relatedness produces its effect. A generalization mechanism may be involved. Thus, recall of a free association stimulus word may automatically increase the availability of other words related to it in meaning. It is possible, however, that the effective mechanism involves mediational strategy. Such a strategy would develop more readily with semantic associates and would provide added information for clustered retrieval. Whatever the mechanism, its effects were not very strong. The overall mean number of semantic clusters was only 1.59 out of a possible 5.67 (half the amount recalled). It seems that when weakly associated experimental pairs are employed mechanisms that promote clustering of experimental pairs must compete with the subjects' tendency to resort to more idiosyncratic pairings to organize their recall. Indeed, Marshall observed substantial amounts of clustering of idiosyncratic pairs. Many such pairs were revealed in this study. Among the more obvious nonexperimental pairs were *baby-box*, *soap-box*, *cold-shoulder*, *tobacco-insect*, *green-tobacco*, and *insect-scared*. On the whole, these pairs did not appear in the free association norms used in this study, so their presence cannot be attributed to their greater preexperimental associative strength. The fact that these nonexperimental pairs were uniformly nonsemantic merely reveals that the words in this study were indeed selected so as to eliminate semantic relationships other than those existing between semantic associates.

REFERENCES

- Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during repeated free

- recall. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 489-495.
- Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.
- Marshall, G. R. Stimulus characteristics contributing to organization in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 364-374.
- Palermo, D. S., & Jenkins, J. J. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.
- Winer, B. J. *Statistical principles in experimental design*. New

York: McGraw-Hill, 1962.

REFERENCE NOTE

- E. A. Bilodeau and D. C. Howell, *Free association norms by discrete and continued methods*. ONR Technical Report, No. 1, 1965. (Contract No. Nonr-475(10), Tulane University.)

(Received for publication December 5, 1974.)