PAPERS FROM THE PARASESSION ON LANGUAGE AND BEHAVIOR CHICAGO LINGUISTIC SOCIETY

MAY 1-2, 1981

EDITED BY

CARRIE S. MASEK
ROBERTA A. HENDRICK
MARY FRANCES MILLER

CHICAGO LINGUISTIC SOCIETY
THE UNIVERSITY OF CHICAGO
CLASSICS 314A
1050 E. 59TH ST.
CHICAGO, ILLINOIS 60637

Relatedness and Strength of Association in Semantic Memory

Carol Jean Godby and Robert A. Fox
Ohio State University

INTRODUCTION. Measures of similarity among words are fundamental to models of semantic processing because they represent an attempt to identify a lower bound necessary for making semantic decisions. According to one conception of similarity widely used in the semantic memory literature, two words are considered to be related if they are members of the same semantic category. For example, sparrows and robins are highly related because both are typical members of the category bird. Accordingly, the distance in memory between sparrow and robin is assumed to be relatively small, and the common properties they exhibit are thought to constitute part of the definition of the abstract category bird.[1] However, there is at least one other plausible conception of similarity. If subjects are asked to name the first word that leaps to mind after hearing chair, they will probably respond with table; chair may also elicit sit or comfortable. The words in this set are highly associated, which also signals proximity in memory, but they are not restricted to membership in the same semantic category.

Just how associative strength and relatedness fit together as measures of similarity has been a matter of controversy in psychological studies of word meaning. Deese (1965), who claimed that the study of association is the most accessible empirical approach to word meaning, argued that associative strength and relatedness are not fundamentally different--or at least that they overlap to a large degree. He came to this conclusion on the basis of subjects' responses to category names like flower. If asked to name as many flowers as possible in three minutes--the essence of a category-naming task--subjects would respond with words like rose and tulip. If asked for free associates to the same word, subjects would respond in more-or-less the same fashion. According to Deese, the similarity of these responses suggests that the category-naming task can be considered a free-association task with an additional stipulation, and that membership in the same semantic category is merely one type of naturally occurring relationship among words. However, others have argued that association and relatedness can be traced to different levels of cognitive organization; in particular, associative strength is thought to be a more primitive measure of similarity than is relatedness.

For example, Anderson (1975) claims that a relation based on category membership such as sparrow and bird can be identified as true or false, but an association like that between salt and pepper is merely strong or weak. Though a semantic relationship may be identified between highly associated words, it is not an essential determinant of associative strength; indeed, semantic content is not necessary, since even nonsense words may be highly associated.

More recently, the issue of associative strength and relatedness has appeared in the research on semantic memory. In the feature-comparison model of Smith, Shoben and Rips (1974) and McCloskey and Glucksberg (1979), the amount of time required to verify a sentence such as all robins are birds is based on the retrieval of features from the nouns in the subject and predicate, upon which a judgment of relatedness is made. If the measure of relatedness is high, as in this example (because a robin is a typical bird), the sentence is verified quickly. However, the same result could have been predicted by using associative strength as the measure of similarity. A comparison of word association norms such as Deese (1965) and category membership norms such as Battig and Montague (1969) reveals that the most typical member of a category is also its most frequent associate. It is only in false sentences that the decision to base similarity judgments on relatedness rather than associative strength leads to different predictions. In a procedure analogous to that of verification, the feature-comparison model predicts that a sentence is falsified quickly when the subject and predicate are in different semantic categories; i.e., if they have few overlapping features, as in all rocks are birds. If the subject and predicate have overlapping characteristics, as in all bats are birds (which might be confused in a speeded reaction time test because both can fly), then response time is inhibited. Note that if associative strength had been the measure of similarity in these examples, there would have been no basis for a prediction of different reaction times, since the nouns in neither sentence are highly associated. Accordingly, no mention is made of the contribution of associative strength to false response times in any of the literature in which the feature-comparison model is developed.

A competing theory of semantic memory is the marker-search model of Holyoak and Glass (1974, 1975). In this model, words in memory are linked in a network, the strength of the links being determined by membership in the same semantic category and associative strength. A sentence is verified by entering the network and searching until the link between the subject and predicate is found. For example, the sentence A robin is a bird is verified quickly because the link between the words is strong and accessible. Holyoak and Glass make the additional claim that the direction of the effect attributed to the strength of the link is the same, regardless of the truth value of the sentence; moreover, the strength of the link is assumed to play a role analogous to that of relatedness in the feature-comparison model. They obtained evidence showing that false sentences exhibiting strong links between subject and predicate, such as all mountains are valleys are rejected more quickly than sentences with less strong links, such as all chairs are beds. This data represents counterevidence to the predictions made by the feature-comparison model.

However, the status of this counterevidence is called into question once the assumption is made explicit that Holyoak and Glass conflate association and relatedness in their measures of the strength of connecting links between words. In the true sentences, such as A robin is a bird, the connection is both highly associated and highly related; among the false sentences, those which elicited the fastest reaction times were rated by subjects in McCloskey and Glucksberg (1979) to be low-related and highly associated. Conversely, the same subjects rated the false sentences which produced the slowest reaction times to be highly related, but low associated. Thus if any single phenomenon is responsible for the same direction of influence in both true and false sentences, it must be associative strength. However, McCloskey and Glucksberg's results suggest that the response times for the false sentences is attributable to relatedness, not associative strength. Note that given McCloskey and Glucksberg's interpretation of Holyoak and Glass's results for false sentences, they no longer represent counterevidence to the feature-comparison model, since as predicted, low-related sentences are falsified quickly, while high-related sentences are falsified more slowly. Thus the assumption is made once again that associative strength has no bearing on semantic decisions.

This hypothesis receives support from the work of Lubker (1979). In his experiments, subjects were timed as they named a picture of a common object with a word superimposed on it. Response time was inhibited if the word and picture were highly related (for example, if the word rat appeared on a picture of a mouse), but there was no inhibition if the word was merely a common associate of the pictured item (for example, if a picture of a mouse was superimposed with the word cheese). Despite this independent evidence from a modified Stroop task that associative strength and relatedness are distinguished with respect to some types of problems, it is necessary to establish direct evidence for this distinction in sentence-verification tasks, since it has been implicated as important for deciding between competing models of semantic processing.

The experiment described below is an attempt to provide such evidence. In the simplest sense, it can be interpreted as a generalization of the paradigm used by Holyoak and Glass, with the exception that association and relatedness are separated. Only false sentences are of theoretical interest, since the predicted effects of both association and relatedness is the same in true sentences. It is a straightforward matter to construct a stimulus set which includes all combinations of associative strength and relatedness:

(1) High associated, high related (HAHR)
    example: All lettuce is cabbage
(2) High associated, low related (HALR)
    example: All mountains are valleys
(3) Low associated, high related (LAHR)
    example: All horses are mules
(4) Low associated, low related (LALR)
    example: All desks are tomatoes

The major theories of semantic memory make the following predictions:

(A) Both associative strength and relatedness facilitate a false response. This hypothesis should predict that LALR is the slowest, unless anomalous sentences introduce a processing lag unrelated to the issues of interest in this test; in such a case, LAHR or HALR should be the slowest. Hypothesis A is consistent with the marker-search model of Holyoak and Glass (1975), although their limited stimulus set does not provide evidence for associative strength and relatedness having an effect in the same direction because it lacks sets of examples like (1) and (4).

(B) Associative strength neither facilitates nor inhibits falsification time, but relatedness inhibits it. This hypothesis predicts that reaction times would divide the stimulus set into two classes: high and low-related. Within each class, there should be no appreciable differences in reaction time. Hypothesis B is compatible with the feature-comparison model.

(C) Associative strength facilitates reaction time, but relatedness inhibits it. This prediction is consistent with the spreading-activation model of Collins and Loftus (1975), which is in many respects a hybrid of the feature-comparison and marker-search models. Like the feature-comparison model, it accounts for the speed of semantic decisions in terms of accumulating positive and negative evidence. Like the marker-search model, it postulates a network representation of semantic memory, in which a variety of strategies other than the comparison of features are employed for reaching semantic decisions. Positive evidence for a decision in the spreading-activation model consists of, among other things, common superordinates and matching features; negative evidence results from the discovery of distinctive properties and the identification of subordinates as mutually exclusive. Keeping in mind that positive evidence slows a false response while negative evidence speeds it up, it is possible to conclude that relatedness counts among the positive evidence for a decision. Associative strength can be considered negative evidence, if it serves to make more explicit the presence of mutually exclusive subordinates or contradictions. This assumption accords with Holyoak and Glass's hypothesis that sentences with strong links between subjects and predicates are falsified quickly because a contradiction can be readily retrieved. It is also consistent with the observation made by investigators of associative meaning that many associated words of the same grammatical category are antonyms.

METHOD:

Subjects: Sixteen OSU students served as volunteers. All were native speakers of English and right-handed. One subject was later eliminated because his mean in the baseline condition exceeded that of the other subjects by 900 msec.

Materials and Stimuli: The design of the sentence verification task required that the noun pairs in the stimulus sentences be matched in terms of low or high relatedness and association. Degree of relatedness was determined by a semantic rating task. Twenty one subjects, none of whom participated in the main experiment, were presented with 65 noun pairs and asked to rate pairwise relatedness on a 7-point scale (7 was low-related). Subjects were explicitly warned about the difference between association and relatedness and told to make their judgments on the basis of relatedness only. Noun pairs with a mean rating of less than 3.5 were considered highly related; those with a mean rating of greater than 5.8 were low-related. The remaining stimuli were constructed by consulting word-association norms in Postman and Keppel (1970) and Deese (1965) to discover noun pairs of equivalent frequency that were disjoint, but of the same semantic category. Each noun selected was the most frequent or second most frequent associate. These pairs were designated as high-associated. The low-associated pairs were selected by their failure to appear in the norms; this assumption was checked by giving them as a free-association test in a separate part of the ratings task. On the basis of these judgments the noun pairs were assigned to one of four conditions: high-associated/high-related; high-associated/low related; low-associated/high-related; and low-associated/low-related. There were a total of 60 stimulus items, 15 in each condition. Half of the sentences were quantified with some, half with all.[2] The test sentences were interspersed among 120 filler sentences, and the entire set of 180 stimuli was randomly ordered, with the constraint that items from each test condition be spread over the entire set. Among the fillers were enough true and false sentences to make a total of 90 true sentences and 90 false sentences.

Procedure: Stimuli were presented on a 2-channel tachistoscope. Subjects indicated their true-false decision by pressing the appropriate response button. Reaction times were measured using a Hunter Klockounter.

Each subject was tested individually. Subjects were told to respond as quickly as possible without sacrificing accuracy. Approximately five seconds elapsed between trials, during which the experimenter recorded reaction times. The entire session lasted one hour, which included time at the end for a discussion of the experiment.

Results: Mean reaction times and error rates for each of the experimental conditions are shown in Table I. They were computed on the basis of correct false responses for 56 of the experimental sentences. Four were eliminated from the analysis because their large error rates indicated that they were inappropriate.

Table I: Mean reaction times and error rates (in parentheses) in each experimental condition.

	LA	HA
LR	1085 (4.1%)	1190 (4.8%)
HR	1294 (7.7%)	1401 (8.4%)

This experiment represented a repeated-measures factorial design having associative strength and relatedness as fixed effects, with words and subjects as random effects (Winer, 1971). Presence of these two random effects in the design required calculation of minimum quasi-F ratios (Clark, 1973) for main effects and their interaction. There was a significant effect for both relatedness (min $F'=12.05$, $df=1/25$, $p<.005$) and association (min $F'=4.74$, $df=1/37$, $p<.05$), but the interaction between associative strength and relatedness was not significant. Error rates in the two high-related conditions were larger than in the low-related conditions.

Given that the LALR condition represents the baseline condition (wherein the effects of association and relatedness are nil), the comparison of reaction-time means in Table I indicates that the effect of high relatedness on reaction times (211 msec) is almost exactly twice that of association (105) msec), and that their combined effects are additive. A t-test demonstrated a significant difference in mean reaction time between the HRLA and the LRHA conditions ($t=2.71$, $df=26$, $p<.01$, one-tailed).

DISCUSSION:

These results, which show that both relatedness and associative strength inhibit falsification, are problematical for all three of the hypotheses considered earlier.

Hypothesis A provides the least satisfying account, since it predicts a direction of influence which is the reverse of that obtained. Moreover, it casts doubt on Holyoak and Glass's (1975) central claim that the discovery of a contradiction is a sufficient strategy for falsifying sentences. If the discovery of a contradiction leads to fast falsification times and associative strength is a measure of the accessibility of the disconfirming superordinate link, then there should be no appreciable differences between the HAHR and the HALR conditions.

Hypothesis B is also not supported, since it predicts that associative strength should have no effects on falsification times. However, this hypothesis is a little more difficult to dispense with because of an alternative explanation concerning the effects attributed to associative strength: these results could be interpreted as meaning that associative strength merely results in more common properties, leading to a greater measure of relatedness. This interpretation receives its strongest support in the two low-related conditions, where the difference in meaningfulness may have produced the reaction-time differences. However, in the high-related conditions, there is no perceptible difference in meaningfulness, yet the absolute contribution of associative strength is the same. The lack of interaction between the two main effects could also be interpreted as indicating that relatedness and associative strength make separate contributions. Nevertheless, this matter is still unresolved and will be the subject of future work.

Hypothesis C provides the most satisfying account of our results at this stage. Since Collins and Loftus (1975) assume that related words have more links with one another than do associated words, activation is predicted to be stronger when related words are involved; however, highly associated words should exhibit a weaker, but still pronounced effect. This is consistent with the results we obtained. The major problem with Hypothesis C is that it predicts an effect in the wrong direction for associative strength. This is a consequence of the assumption that the contribution of associative strength is negative evidence for a semantic decision (because of its presumed role in identifying items as contradictory), which has the effect of speeding response time for false sentences. Whatever the contribution of associative strength may turn out to be, the success of Hypothesis C requires it to be counted in the list of positive evidence.

The goal of this work has been to clarify a terminological confusion which runs through the semantic memory literature by making as explicit as possible with current empirical procedures the distinction between association and relatedness. It is certainly encouraging that even with the more restricted definition of psychological similarity that relatedness is identified as overwhelmingly more important than association for making decisions of the type studied so far in semantic memory research. Our results may thus be interpreted as suppost for a distinction between associative and semantic memory, with the measure of similarity originating from the deeper level of cognitive organization being most relevant for semantic decisions. However, the contribution of associative strength cannot be ignored. Identifying its precise role in the falsification of sentences will provide future tests for the sensitivity of semantic memory models.

[1]However, it is simplifying matters somewhat to say that relatedness is based solely on membership in the same semantic category. Two additional points must be made:
(i) Relatedness ratings are obtained from the results of multidimensional scaling tests (on which a subject is asked, for example, to determine the similarity of door and window on a 7-point scale), not the intuitive judgments of the experimenter. Membership in the same semantic category is one salient dimension extracted from such results, although subjects may be using other criteria for making their decisions. See Smith, Shoben and Rips (1974) for discussion.
(ii) Membership in the same semantic category is in itself an insufficient measure of relatedness because there are at least two ways in which it can be involved:
(a) The distance between the category name and a particular instance is relatively small--a Euclidean account of the typicality effect. For example, car and vehicle are highly related because car is a typical vehicle; skateboard and vehicle are considerably less related. (cf. Rosch 1975)
(b) The distance between instances of the same category is small. Because we used natural-kind terms in our experiments, we interpret a high similarity rating (as obtained, for example, between moth and butterfly) as meaning that the words in question share many salient perceptual features, or are most likely to be confused with one another.

[2]Both _all_ and _some_ were used in the construction of the stimulus sentences in part to obscure the nature of the task. This procedure did not affect the outcome of the experiment, since _t_-tests showed no significant differences in quantifier types, a result consistent with Glass and Holyoak (1974); Glass, Holyoak and O'Dell (1974); and Holyoak and Glass (1975).

References

Anderson, B. _Cognitive Psychology_. New York: Academic Press, 1975.

Battig, W.F. and Montague, W.E. Category norms for verbal items in 56 categories: a replication and extension of the Connecticut category norms. _Journal of Experimental Psychology Monograph_, 1969, _80_.

Clark, H.H. The language-as-fixed-effect-fallacy: a critique of language statistics in psychological research. _Journal of Verbal Learning and Verbal Behavior_, 1973, _12_, 335-359.

Collins, A.M and Loftus, E.F. A spreading activation theory of semantic processing. _Psychological Review_, 1975, _82_, 407-428.

Deese, J. _The Structure of Associations in Language and Thought_. Baltimore: Johns Hopkins University Press, 1965.

Glass, A.L and Holyoak, K.J. The effects of _some_ and _all_ on reaction time for semantic decisions. _Memory and Cognition_, 1974, _2_, 436-440.

Glass, A.L., Holyoak, K.J. and O'Dell, C. Production frequency and the verification of quantified statements. _Journal of Verbal Learning and Verbal Behavior_, 1974, _13_, 237-254.

Holyoak, K.J. and Glass, A.L. The role of contradictions and counterexamples in the rejection of false sentences. _Journal of Verbal Learning and Verbal Behavior_, 1975, _14_, 215-239.

Lupker, S.J. The semantic nature of response competition in the picture-word interference task. _Memory and Cognition_, 1979, _7_:6, 485-495.

McCloskey, M. and Glucksberg, S. Decision processes in verifying category membership statements: implications for models of semantic memory. _Cognitive Psychology_, 1979, 11, 1-37.

Postman, L. and Keppel, G. (Eds.). *Norms of Word Association*. New York: Academic Press, 1970.

Rosch, E.H. On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press, 1973. Pp. 111-144.

Smith, E.E., Shoben, E.J. and Rips, L.J. Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214-241.

Winer, B.J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1971.