Journal of Experimental Psychology
1973, Vol. 100, No. 2, 429–431

SEMANTIC MEMORY AND SENTENCE VERIFICATION TIME[1]

THEODORE J. DOLL,[2] JAMES R. TWEEDY, MARCIA K. JOHNSON,
JOHN D. BRANSFORD, AND CARL FLATOW

State University of New York at Stony Brook

In 1969, Collins and Quillian adduced evidence that semantic memory is hierarchically organized by comparing verification times for sentences in which the subject and predicate are more or less remote in a logical hierarchy. However, they confounded particular sentence subjects with the distance separating the subject and predicate. The present study resolved the confounding by using sentences which had the same subjects, but had predicates presumably stored at different levels in the hierarchy. This study also tested the generality of the hierarchical model by using slightly more complex sentences. The results provide some support for Collins and Quillian's model, both in the case of simple and more complex sentences. Problems in generalizing from these results and limitations of the hierarchical model are discussed.

Quillian (1968, 1969) and Collins and Quillian (1969, 1970) have proposed a hierarchical model of semantic memory organization. One prediction of this model is that verifying a statement like *A canary can fly* should take longer than verifying a statement like *A canary can sing*. This follows from the assumption that the property *can fly* is further removed in the hierarchy from *canary* than is the property *can sing*. Collins and Quillian (1969) obtained reaction time (RT) results consistent with this prediction. However, an inspection of their materials yields possible alternative explanations of their results.

One possible account of the Collins and Quillian (1969) data involves bias in terms of the subject nouns used in their sentences. Their lists were constructed so that the basic comparison was of sentences like *A canary can sing* and *Champagne is sparkling* (lower level properties) with sentences like *A wren can fly* and *Chianti is alcoholic* (higher level properties). Sentence subjects occurred only with properties from one level and were not balanced across property levels. As the above examples indicate, sentences in which the property is stored at the lowest level or node of the presumed hierarchy must have subject nouns that are generally known and that have highly salient distinguishing properties. It may be that, in general, *Ss* can gain access to information about such nouns faster than they can gain access to nouns with less salient distinguishing properties, e.g., those used in sentences with the higher level properties.

A study reported by Conrad (1972) suggests another possible source of bias in the Collins and Quillian (1969) materials. Conrad conducted 2 experiments varying both subject–property remoteness and the frequency with which properties were given as descriptive responses to their corresponding node items (e.g., *can sing* as a response to *canary* and *can fly* as a response to *bird*). In Experiment I, the remoteness effect seemed to be limited to certain levels of the frequency variable. However, this experiment suffered from the same difficulty as Collins and Quillian's (1969) study in in that subject nouns were confounded with the independent variables. Conrad tried to overcome this problem in a second experiment by presenting the sentence subject 1 sec. in advance of the property and timing RT from the onset of the property. As Conrad points out, this procedure allows *S* to encode the sentence subject before the timing of RT begins. However, it may also allow *S* to do additional processing (such as forming hypotheses about the property) which could obscure any effect of the remoteness variable. This seems especially likely since the properties in her experiment were always drawn from the same level of the hierarchy (e.g., if the subject was any sort of animal, the property was always a characteristic of animals in general) and *S* apparently saw some properties more than once in the same session. Therefore, her procedure for controlling for possible differences in time to gain access to subject nouns may not be the most appropriate.

The present study eliminated the confounding of sentence subject with remoteness more directly, by comparing verification times for sentences which had the same subjects but which had predicates stored at different levels in the hierarchy, e.g., *Elephants have trunks* vs. *Elephants have skin*.

A secondary purpose of the present study was to test the generality of the hierarchical model by using sentences which contained a predicate adjective (e.g., *Elephants have long trunks* and *Elephants have tough skin*). Adjectives were chosen so that they were particularly appropriate to the property with which they appeared and could not be applied to all properties of the subject or its

[1] This research was supported in part by grants from the Joint Awards Committee/University Awards Committee of the State University of New York Foundation, and the Graduate School and Institute for Research in Learning and Instruction of the State University of New York at Stony Brook. The authors are grateful to Edward Lambe, Director of the Instructional Resources Center at the State University of New York at Stony Brook, for making available the Computer Assisted Instruction Laboratory for this research.

[2] Requests for reprints should be sent to Theodore J. Doll, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790.

TABLE 1

EXAMPLE SENTENCES AND RESULTS

Type	Remoteness[a]	Example sentence	Correct response	Mean reaction time (in msec.)	Error (%)
1	0	Elephants have trunks	true	1,629	2.19
2	≥ 1	Elephants have skin	true	1,679	4.16
3	0	Elephants have long trunks	true	1,873	5.47
4	≥ 1	Elephants have tough skin	true	1,925	6.17
5	—	Elephants have rudders	false	1,815	3.40
6	—	Elephants have diesel trunks	false	2,150	5.60
7	—	Elephants have long rudders	false	2,216	3.57
8	—	Elephants have diesel skin	false	2,116	3.89
9	—	Elephants have tough rudders	false	2,143	2.72

[a] Remoteness = number of nodes separating subject and predicate in hierarchical model proposed by Collins and Quillian (1969).

superordinates. The question was whether verification times would still reflect hierarchical organization when the sentences to be processed were somewhat more complex than those used in previous studies.

Method. Ninety sets of 9 sentences each were used. A typical set is shown in Table 1. All the sentences in a given set had the same subject and verb. Eighty-one sets were made up of sentences which specified property relationships, e.g., *Elephants have skin*; 9 sets contained sentences which specified categorical membership, e.g., *Rats are animals.* Each set was begun by generating 2 true sentences in which the predicate could be correctly applied only to the subject of the sentence (Type 1), or to the subject and some superordinate category of which the subject was an instance (Type 2). For each set, 2 additional true sentences, Types 3 and 4, were generated by adding appropriate modifying adjectives to the Type 1 and Type 2 sentences. False sentences (Types 5–9) were produced by substituting an inappropriate noun or inappropriate adjective in the predicate, as illustrated in Table 1.

Nine sequences of sentences were generated by selecting one sentence from each of the 90 sets. Each sequence included 10 instances of each of the 9 sentence types. Each *S* was assigned to one of the 9 sequences, so that each *S* saw only one sentence from each of the 90 sets.

Forty students from a learning course at the State University of New York at Stony Brook participated in the experiment. They were run in 2 groups of 20 each, and were paid $1.50 for the session, which lasted approximately 1 hr.

Each *S* was seated in front of a terminal which consisted of a cathode ray tube (CRT) display and a typewriter keyboard. The experiment was controlled by an IBM 1500 computer, which automatically recorded responses and measured RT to the nearest .1 sec. Subjects were instructed to press the *Z* key for each true and the / key for each false sentence as it appeared on the CRT. They were instructed to make their responses as quickly as possible without making errors, and to avoid making metaphoric or poetic interpretations of the sentences. The experiment began with a block of 30 practice sentences. The sequence of 90 experimental sentences was divided into 3 blocks of 30 sentences each.

The *S* was given 4 sec. to respond to each sentence before a warning signal for the next trial appeared. The delay between a response and the presentation of the next sentence was about 4 sec. A 30-sec. rest period was provided between each block of sentences.

Results and discussion. Mean RTs for the 9 sentence types were first computed separately for each individual *S*. Inspection of the RT data revealed no differences as a function of sequence; hence the data were pooled over sequences for subsequent analyses.

Mean RTs for correct responses and error rates are shown in Table 1. Since the error rates are all very low, RT differences can safely be attributed to differences in sentence processing time rather than to guessing or speed–accuracy trade-off effects.

Separate analyses of variance were performed for true and false sentences. The most important finding was that mean RT for true sentences increased with the remoteness of the subject and predicate in the presumed hierarchy, $F\ (1, 39) = 4.15$, $p < .05$. The increase in RT with remoteness averaged 50 msec. for sentences without adjectives and 52 msec. for sentences with adjectives. Obviously then, the interaction between remoteness and the presence or absence of an adjective was not significant. Of course, sentences with adjectives had longer RTs than sentences without adjectives, $F\ (1, 39) = 129.09$, $p < .01$. This finding was expected on the basis that sentences with adjectives should take slightly longer to read.

For false sentences, linear contrasts were used to evaluate certain comparisons. Mean RT for sentences without adjectives (Type 5) was significantly faster than mean RT for sentences with adjectives (Types 6–9), $F\ (1, 39) = 91.06$, $p < .01$. Another contrast compared sentences containing inappropriate adjectives (Types 6 and 8) with sentences with appropriate adjectives but inappropriate predicate nouns (Types 7 and 9). Although mean RT was slightly higher for Types 7 and 9, the differences did not approach significance ($F < 1$). This result suggests that sentences are probably not verified simply by reading from left to right and terminating when a falsifying word is encountered.

It was also of interest to determine whether the remoteness effect still obtained when sentences that specified properties were considered independently of those that specified categorical relationships. Mean RTs for correct responses to property sentences, Types 1–4, were, respectively, 1,623, 1,696, 1,882, and 1,933 msec. Thus the remoteness effect seems to have been at least as strong in property sentences as in the entire set of experimental sentences.

Conclusion. The present study indicates that the subject–predicate remoteness effect reported by Collins and Quillian (1969) is still obtained when (*a*) the test sentences representing different degrees of remoteness refer to the same sentence subjects, and (*b*) when the sentences are somewhat more complex than in previous studies, i.e., when a predicate adjective is added. However, further considerations suggest some problem in generalizing from the above findings. For one, the authors can testify to the difficulty of generating 90 sets which, on logical grounds alone, can be regarded as hierarchically organized (e.g., under what superordinate is *telephone* classified?). More important, in all studies relevant to the hierarchical model, the remoteness variable has probably been confounded with a number of other factors. For example, Conrad's (1972) results suggest that frequency of properties as responses to node items may contribute to sentence verification RT. Jorgensen and Kintsch (1973) have reported another possible source of variation. They found that sentence imagery value affects verification RT.

It seems likely that a strictly hierarchical model is insufficiently general to serve as a complete characterization of the organization of semantic memory. Collins and Quillian have themselves pointed out that "hierarchies are not always clearly ordered [1969, p. 242]" and "people surely store some properties at more than one level in the hierarchy [1969, p. 242]." Therefore, a fruitful strategy for future research might be to try to characterize other principles of organization which complement the hierarchical principle in determining the organization of semantic memory.

REFERENCES

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

Collins, A. M., & Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1970.

Conrad, C. Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 1972, **92**, 149–154.

Jorgensen, C. C., & Kintsch, W. The role of imagery in the evaluation of sentences. *Cognitive Psychology*, 1973, **4**, 110–116.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: M.I.T. Press, 1968.

Quillian, M. R. The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, 1969, **12**, 459–476.

(Received September 28, 1972)