AMERICAN JOURNAL OF PSYCHOLOGY
June 1979, Vol. 92, No. 2, pp. 363-375

0002-9556/79/0600-0363$01.30/0

Integration of verbal and visual information as evidenced by distortions in picture memory

Dedre Gentner
University of Washington

Elizabeth F. Loftus
University of Washington

Subjects were presented with a series of pictures, some of which were general (girl walking down the path) and others specific (girl hiking down the path). These pictures were matched with sentences which were either general or specific ("The girl is walking [hiking] down the path.") Subsequently, a forced-choice picture recognition test was administered in which subjects saw pairs of pictures and indicated which member of each pair they had seen before. It was found that labelling the picture with a sentence containing a specific verb substantially increased the likelihood that the specific picture corresponding to that verb would subsequently be falsely recognized. The results are discussed in terms of current theories of memorial representation.

How is semantic information from different modalities integrated and stored? If related ideas are encountered in pictures and sentences, is the result a single representation in memory or two representations that separately hold information from the two modalities?

These questions were raised by Rosenberg and Simon (1977) in a study concerning the effects of presenting semantic information in different modalities. Rosenberg and Simon presented subjects with pictures and sentences that expressed related information. In a later recognition test, some of the previously seen sentences were expressed as pictures, and some of the pictures were expressed as sentences. Subjects frequently falsely recognized these translations as familiar; that is, they had difficulty discriminating translations from previously presented material. These results were interpreted to mean that presenting material in different modes may produce identical representations. If a new sentence or picture matches the meaning of an underlying preexisting representation, it will appear to be familiar and result in a false recognition.

This interpretation is consistent with the work of Pezdek (1977) as well. In her research, subjects were presented with a sequence of

pictures and sentences, then presented with intervening pictures and sentences, and finally given a recognition test. When an intervening item (such as *eagle*) was semantically relevant to an earlier item (such as a picture of a bird), yes-no recognition accuracy on the earlier item was reduced relative to the case when a semantically irrelevant item intervened. This effect occurred despite the fact that the intervening items were presented in a different modality from the earlier to-be-recognized items. These results were taken to support the notion that information from different modalities can be integrated in memory.

These findings might be considered damaging to dual-storage models of memory, which posit that verbal information and visual information are represented in separate stores (cf. Paivio, 1971). For if recognition is assumed to be a process of matching stimuli to modality-specific representations in a dual-modality storage system, then a previously presented picture or sentence should result in a match, whereas a translation from one modality into the other or an integration between the two modalities should fail to match. However, it could be argued that the yes-no recognition procedure used by Pezdek and by Rosenberg and Simon is not sensitive enough to reveal all that the subject knows. A subject given a test stimulus that is a translation of an original stimulus into another modality may notice the correspondence of content between the test stimulus and the original-modality memorial representation. In the absence of a better match, this correspondence might lead him or her to a false recognition of the test stimulus. Thus, poor performance on such a yes-no recognition task does not necessarily constitute evidence against a dual-storage theory. However, a forced-choice procedure should reveal more about the subject's knowledge. In a forced-choice procedure, a subject would be given a choice between a translation of an item into another modality and the original item itself. If modality-specific memorial representations exist, subjects should be able to make the discrimination between the precise stimulus they have seen and a translation from another modality.

An example of the sensitivity of the forced-choice procedure is a study done by Anderson and Bower (1973) as a follow-up to a highly influential study by Bransford and Franks (1971). Bransford and Franks presented subjects with a series of simple sentences derived from one complex scenario. In a later yes-no recognition test, it was discovered that subjects were unable to discriminate old sentences from new sentences also derived from the scenario. This was taken to indicate that the simple sentences had been integrated into a single complete representation. Anderson and Bower criticized the proce-

dure, arguing that the yes-no recognition test was too insensitive to support the conclusions; more specifically, they claimed that presenting one test sentence at a time does not force the subjects to use whatever specific information they might possess. When a forced-choice recognition procedure was instituted, subjects showed some ability to discriminate between old and new sentences, particularly when those sentences contained relatively few propositions.

An additional problem that arises in any recognition procedure designed to measure modifications in memory is that since a subject's memory alterations are idiosyncratic, the precise nature of the distractor stimuli is crucial. Suppose a subject sees a picture of a woman walking and in some way integrates the memory of the picture with the verbal statement that the woman was hiking. Later, the person is shown a picture of a woman *hiking*—perhaps wearing a backpack and boots. Will the subject falsely recognize the picture? Only if his or her memory modification agrees with the particular modified stimulus that is presented. Otherwise, the subject might reject the new stimulus despite actual modifications in memory. There is thus a danger of underestimating the amount of change that takes place in memory.

The more accurately we can anticipate the kinds of changes that will take place in a person's memory, the better we can test whether memory modification has in fact taken place. However, our ability to accurately predict these modifications in memory depends upon our possessing clear models of how knowledge is represented in memory. One area in which reasonably well-specified models of meaning have been developed is that of verb meaning. The many models differ from one another in detail, but most assume an analytic representation based on semantic components or underlying predicates (e.g., Gentner, 1975, 1977; Lakoff, 1970; McCawley, 1968; Rumelhart & Levin, 1975; Schank, 1972; Talmy, 1975). In these analytical models, the process of integrating information from two sources can be characterized as combining semantic components into a single representation. In one study, Gentner (1975) utilized pairs of general/specific verbs, whose representations have a fairly clear relationship. The representation of the specific verb contains the entire representation of the more general verb, as well as additional semantic information. For example, *pay* consists of all that is contained in *give*, plus additional components. Gentner showed that subjects who heard the general verbs in the context of the additional components falsely recalled hearing the specific verb, suggesting that they had integrated the information into a single representation. Our approach to cross-modal integration makes use of these notions. We assume that when a

subject views a picture, for example, of someone walking, some representation is formed in memory. If the person now labels the picture with a specific verb, such as *hiking*, additional semantic information will be added to the original representation. The person's memorial representation will then more closely match a more specific picture containing the new information than it will the original picture. Consequently, false recognitions will occur, even in a forced-choice recognition test.

These ideas led to the present experiment, which was designed to test for specifically predicted cross-modal integrations. Subjects were presented with a series of general or specific pictures, each of which was matched with a sentence that was either general or specific. For example, subjects would see a picture showing a girl either walking (general) or hiking (specific) down a path (see Figure 1). The picture was matched either with the sentence, "The girl is walking down the path" (general), or with the sentence, "The girl is hiking down the path" (specific). The experimental question was whether subjects' memory for the picture would be altered by the kind of sentence presented. More precisely, the question was whether specific information presented in a sentence would be added to the subject's representation of the picture to produce a more specific picture representation. If the specific verb concept, such as hiking, is integrated into the representation of the general picture (the girl walking), then the result will be a composite. In that case a picture of a more specific event (e.g., the girl hiking) may match the altered representation more closely than does the original picture. False selections of the more specific picture will result. Thus the major hypothesis of our experiment was that the subjects would falsely recognize the specific pictures more often when sentences containing specific rather than general verbs were used to label general pictures.

METHOD

Subjects

The subjects were 36 students enrolled in psychology courses at the University of Washington. They received course credit for their participation. Subjects were run in small groups.

Materials

The experimental material consisted of 16 critical pairs of pictures and 16 critical pairs of sentences. The members of each pair of sentences were identical except that one sentence contained a general verb and the other contained a more specific verb. For example, the sentence "The girl is walking

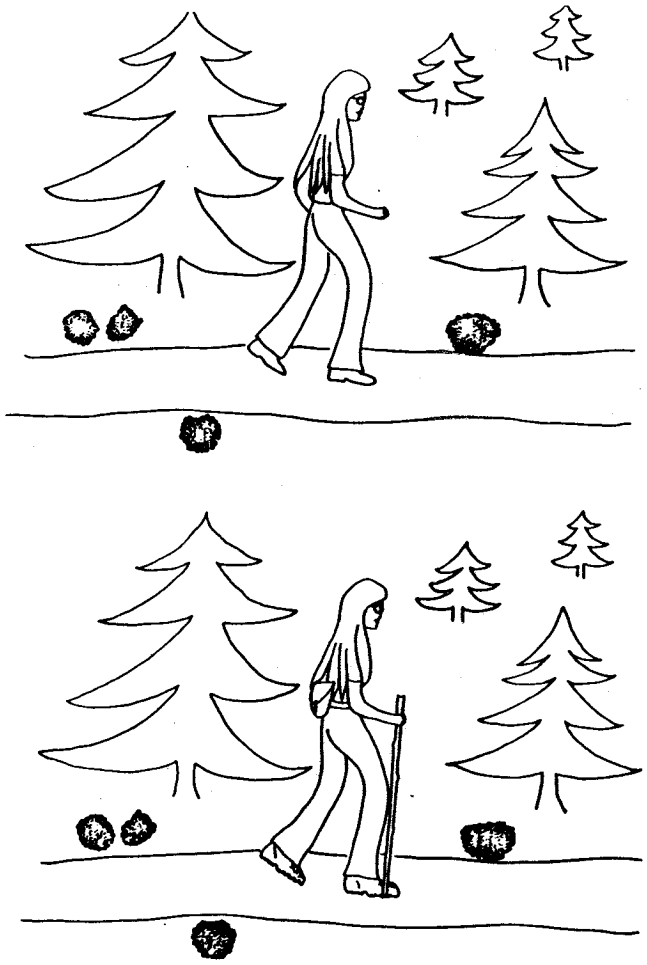

Figure 1. General picture (top) shows girl walking down a path; specific picture (bottom) shows girl hiking down a path

down the path" contains the general verb *to walk*, while "The girl is hiking down the path" contains a more specific verb, *to hike*. Each general/specific sentence pair corresponded to a similar pair of pictures. In this example, the general picture shows a girl walking; the specific picture is identical, except that the girl has a pack and boots and is carrying a staff. Figure 1 presents this pair of pictures. The entire list of general and specific verb pairs is given in Table 1.

Table 1. Verbs used in the experiment

General verb	Specific verb
Break	Smash
Cook	Bake
Eat	Picnic
Get	Buy
Greet	Hug
Hurt	Burn
Sit	Slouch
Smile	Laugh
Split	Cut
Talk	Argue
Tend	Pick
Walk	Hike
Watch	Time
Work on	Paint

Design and procedure

Each group of subjects was presented with a random ordering of 16 critical slides. Only one member of a pair was seen by any given subject. There were also 16 filler slides; two of these preceded and two followed the critical slides, and an additional 12 were interspersed among the critical slides. Each slide appeared for two sec.

Each subject had a sheet of paper containing a list of 39 sentences. The subjects matched each slide as it appeared to one of the sentences on the sheet. There were 16 critical sentences that matched the critical slides, 16 filler sentences that matched the 16 filler slides, and an additional seven unmatched filler sentences. Only one member of a pair of sentences was seen by any given subject. After each slide was presented, the subject indicated which sentence best described the slide by marking the slide number next to that sentence. This matching process was self-paced, with a median time of 25 sec for a group to match. The experiment was designed so that each picture and each sentence appeared in its general and its specific form equally often across subjects, and so that each subject saw an equal number of specific and general pictures and sentences.

The stimuli were designed with no close distractors so that the critical sentence-picture pairs could readily be matched, regardless of the generality/specificity match. For example, in the walking/hiking pair, the general picture of the girl walking could be matched either with the general sentence, "The girl is walking down the path," or with the specific sentence, "The girl is hiking down the path." The same was true when the specific picture was presented. There were no other plausible matches. As expected, subjects were very accurate at the sentence-picture matching. The data from two subjects who made more than two matching errors were discarded and these subjects were replaced with two new subjects. All subjects were asked to return one week later but were given no reason for doing so. (Two other tasks,

not relevant here, were also included in the initial session; we felt that it would be difficult for subjects to guess the reason for the return session.)

One week after completing the picture labelling task, subjects were given a forced-choice recognition test on pairs of pictures. The purpose of the delay was to allow for the waning of surface memory (since the controversy of dual versus single storage concerns long-term memory). The test consisted of showing the 16 critical pairs of pictures and requiring the subject to indicate which member of each pair he had seen before. The correct picture appeared equally often in the left and right position. Picture pairs were shown for one min.

RESULTS

Table 2 shows the proportions of times subjects chose the specific picture in each of four picture-sentence conditions. Not surprisingly, subjects chose a specific picture more often when they saw a specific picture than when they saw a general picture (72.5% versus 36.5%). They also chose the specific picture more often when the matching sentence contained a specific verb rather than a general verb (62% versus 47%). If we look only within the two conditions in which subjects actually saw the general picture, we find that for pictures labelled with a general verb, there were 27% specific responses, while for pictures labelled with the more specific verb, there were 46% specific responses.

Using procedures outlined by Clark (1973), it was determined that picture type and sentence type both significantly influenced the likelihood that subjects would choose a specific picture. The interaction between these two factors was not significant. The complete analyses are presented in Table 3.

The effect of hearing the specific sentence appears equally strong for general (G) and specific (S) pictures. This was reflected in a lack of interaction between the factors of picture specificity and sentence specificity. The more specific information subjects get, from whatever source, the more likely it is that their representations match the specific picture.

Effects of variation in the appropriateness of the matches

The appropriateness of the picture-sentence matches varied somewhat. In addition to unsystematic variability in the appropriateness of the picture-word pairs, there was the systematic difference that the G/S and S/G matches were of necessity less close than the G/G and S/S matches. Such differences might have affected the way in which the stimuli were remembered and could possibly have biased the results.

Table 2. Proportion of times a specific picture was chosen in a forced-choice recognition test

Type of picture seen	Type of sentence: General	Type of sentence: Specific
General	.27	.46
Specific	.67	.78

Table 3. Analysis of proportions of specific responses

	Subject analysis		Item analysis		Min F	
Source of variance	F	error	F	error	F	p
Pictures	74.41	.99	22.79	6.45	17.45	$<.01$
Sentences	8.52	1.37	10.62	2.72	4.73	$<.05$
Interaction	6.33	.09	1.57	2.88	1.26	$>.10$

For example, any tendency of subjects given a mismatch to discount the more general stimulus could have led to bias in favor of the results obtained. Even a systematic tendency to rely more on either sentences or pictures could have led to bias if the average goodness-of-match had varied across the G/G, G/S, S/G and S/S groups. To investigate these possible artifacts, two groups of 12 subjects each were asked to rate, on a scale of 1 (very poor match) to 5 (very good match), how well the picture-sentences pairs matched. One group rated G/G and S/S sentences; the other, G/S and S/G sentences. This method meant that a subject saw each picture and each sentence once. The mean ratings were as follows: G/G, 3.7; S/S, 3.9; G/S, 2.7; S/G, 3.4. As expected, G/G and S/S pairs are rated as better matches than G/S and S/G pairs. For each of the four kinds of pairs, a Spearman rank-order correlation was calculated between the mean goodness ratings of the 16 picture-sentence pairs and the mean number of times the original subjects had chosen specific pictures in the recognition task. All four correlations were nonsignificant. The *rhos* were as follows: G/G, .246; S/S, −.116; G/S, −.115; and S/G, .184. ($rho^* = .425$ for $p < .05$, $N = 16$). Thus, the recognition results cannot be attributed to differences in the appropriateness of picture-sentence matches.

Adding to versus altering a picture

One final observation is of interest. The transformation of pictures from their general to their specific form was mainly accomplished by

the addition of a few objects. Thus, the picture of the girl walking became a picture of the girl hiking by the addition of a pack, boots, and staff. However, some pairs required some alteration in the general picture to create the specific picture. The pictures shown in Figure 2 are an example of this type. The general picture shows a man sitting in a chair, while the specific version shows the same man slouching in a chair. We saw no evidence that the alteration-requiring pairs behaved differently from the additive pairs. However, the number of alteration-requiring pairs was too low to allow any firm conclusions.

DISCUSSION

The present experiment provides additional confirmation that subjects integrate in memory pieces of information that are related in meaning. The integration process results in a later inability to accurately recall or recognize information as it was originally presented. Subjects who verbally received specific information were more likely to falsely recognize specific pictures than subjects who had not received the specific verbal information. In this instance, the effect of the specific sentence was equally strong for general and specific pictures. Apparently, the more specific information a subject received, from whatever source, the more likely it was that his or her memorial representation matched the specific picture.

Historically, integration phenomena closely related to those studied here were reported by the Gestalt psychologists. In the classic experiment by Carmichael, Hogan, and Walter (1932), subjects were shown line drawings either by themselves or with one of two kinds of verbal labels. Later, they were asked to reproduce the drawings. Their later reproductions were distorted in the direction suggested by the labels. That subjects altered their representations of the figures to correspond with the verbal information is in accord with the results presented here. This sort of memory modification has been observed in recognition tests as well; however, integration phenomenon have been more difficult to demonstrate in recognition than in recall because of the stimulus design problem mentioned in the introduction (Daniel, 1972; Prentice, 1954).

These integration results bear on the issue of how information is represented in memory, whether in one unified store or in two or more modality-specific stores. There are many versions of this dichotomy, but the major issues seem to be as follows: One-store models conceive of memory as one unified-storage system, typically with a propositional representational format (Anderson & Bower,

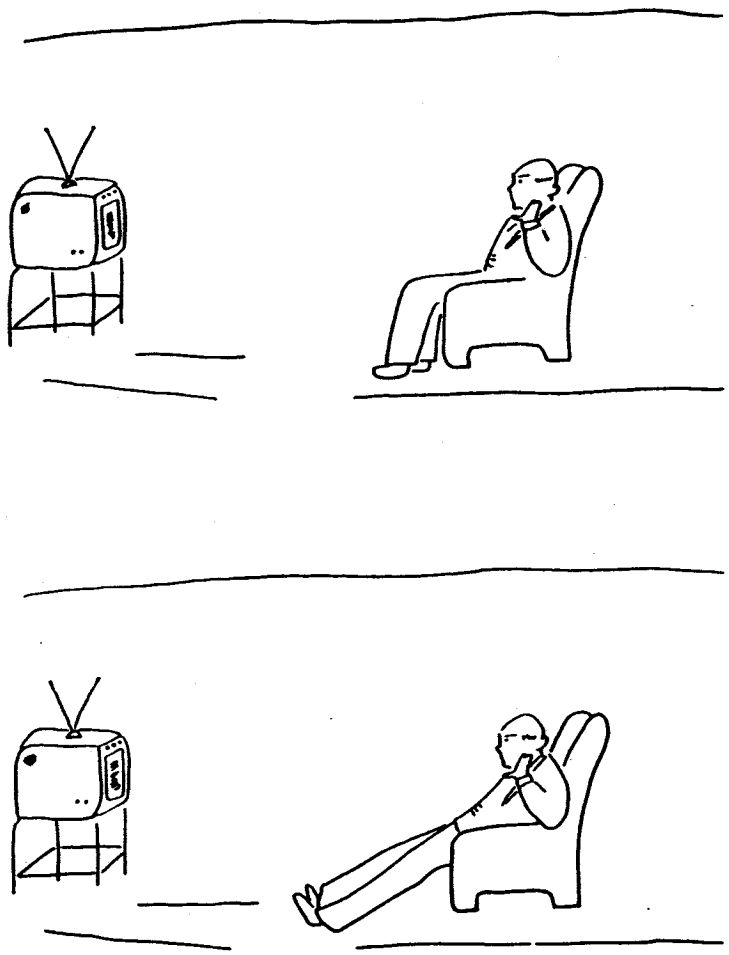

Figure 2. General picture (top) shows man sitting in a chair watching TV; specific picture (bottom) shows man slouching in a chair

1973; Kintsch, 1974; Rumelhart, Lindsay, & Norman, 1972). Here "propositional" should not be taken to mean "verbal"; rather, it means: consisting of conceptual elements representing entities and relationships between entities. (See Palmer (1978) for a more complete discussion.) Dual-storage models postulate a modality-specific memory store for visual information and a separate storage system that is verbal or propositional (Kosslyn, 1975; Paivio, 1971; Shepard & Chipman, 1971). In these models, information in a picture would be stored separately and represented differently from information pre-

sented in a sentence. Exactly how visual information is represented in the visual store is not clear. Kosslyn and Pomerantz (1977) rightly point out that the "pictures-in-the-head" version of the visual-storage model is a bit of a straw man; nevertheless, something on this order seems to be at least the intuitive base for most such models.

Let us now consider the implications of the present results for these two kinds of models. If subjects stored the pictures in a separate store from the sentences, as in the dual-storage model, then some kind of interaction between the two stores must have occurred to produce the integrations observed here. Further, this interaction must not have been under the voluntary control of the subjects, since the task requirement of accurate picture recognition could best have been accomplished simply by utilizing the contents of the visual store and ignoring the contents of the verbal store. Thus, the dual-storage model must be modified to include a considerable amount of interaction between the two stores in order to account for the present results.

The unified-storage model seems to accommodate the integration result more readily. However, here too some circumspection is necessary. Consider the subjects who received general pictures and specific sentences. In the simplest case of the unified model, these subjects should have added the specific information in the sentences to their pictorial representation. They should thus have chosen the specific pictures almost invariably. This, of course, did not happen; rather, subjects were often able accurately to reject the specific picture in recognition. Some of this accuracy is probably attributable to the problem of idiosyncratic modifications discussed above. Yet it seems that provision must be made in the unified-storage model for some memory for modality. Thus, to accommodate the present results, we can either alter the dual-storage model to allow for more interaction between the two stores, or we can postulate a unified-storage model with some modality-specific memory storage. It is not clear that these altered models are really different from one another in any fundamental way. These results and other integration results force us to consider whether the dichotomy between dual storage and unified storage is a useful one.

There are many specialized subsystems that arise in the representation of knowledge. Within the verbal domain itself, we can distinguish subsystems for lexical, syntactic, semantic, and perhaps pragmatic information. Simulations of human language use have found it convenient to treat these systems as semi-autonomous—partially separate but highly interactive (Nash-Webber, 1975; Reddy, Erman, Fennell, & Neely, 1973; Winograd, 1972). We can find other such specialized

subsystems; for example, in modelling human understanding of electronic circuitry, knowledge of procedures and knowledge of enduring facts have been conceived of as separate interactive systems (Brown & Burton, 1975). Perhaps we should begin to consider the general issue of how specialized subsystems are represented in memory and how they interact, rather than focusing on the visual-verbal dichotomy as unique.

Note

The authors thank Mary Kay Riddell for her assistance at all phases of running the experiment. Colin MacLeod and David Marks generously gave their time to discuss the experiment and theoretical issues with us. Dr. Gentner is now at Bolt, Beranek & Newman, Inc., in Cambridge, Massachusetts. The research was supported, in part, by grant MH-27065 from the National Institute of Mental Health to E. Loftus. The writing of the manuscript was facilitated by support from the National Science Foundation, Grants BNS 77-26856 and BNS 76-22943 A 02, and support from the Andrew W. Mellon Foundation. Please send all correspondence and requests for reprints to E. Loftus, Center for Advanced Study in the Behavioral Sciences, Stanford, CA 94305 (before September, 1979), and to E. Loftus, Department of Psychology, University of Washington, Seattle, WA 98195 (after September, 1979). Received for publication August 8, 1977; revision received April 18, 1978.

References

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: Winston, 1973.

Bransford, J. T., & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology,* 1971, *2,* 331–350.

Brown, J. S., & Burton, R. R. Multiple representations of knowledge for tutorial reasoning. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding.* New York: Academic Press, 1975.

Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study on the effect of language on the reproduction of visually perceived forms. *Journal of Experimental Psychology,* 1932, *15,* 73–86.

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.

Daniel, T. C. Nature of the effect of verbal labels on recognition memory for form. *Journal of Experimental Psychology,* 1972, *96,* 152–157.

Gentner, D. Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition.* San Francisco: Freeman, 1975.

Gentner, D. *On relational meaning: The acquisition of verb meaning.* CSR Technical Report No. 78, December 1977.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341–371.

Kosslyn, S. M. & Pomerantz, J. R. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 1977, *9*, 52–76.

Lakoff, G. *Irregularity and syntax*. New York: Holt, 1970.

McCawley, J. D. The role of semantics in a grammar. In E. Bach & R. T. Harms, (Eds.), *Universals in linguistic theory*. New York: Holt, 1968.

Nash-Webber, B. The role of semantics in automatic speech understanding. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Paivio, A. *Imagery and Verbal Processes*. New York: Holt, Rinehart, Winston, 1971.

Palmer, S. E. Fundamental aspects of cognitive representation. In E. H. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Potomac, Md.: Erlbaum, 1978.

Pezdek, K. Cross-modality semantic integration of sentence and picture memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 515–524.

Prentice, W. C. H. Visual recognition of verbally labelled figures. *American Journal of Psychology*, 1954, *67*, 315–320.

Pylyshyn, A. W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.

Reddy, D. R., Erman, L. C., Fennell, R. D., & Neely, R. B. The Hearsay speech understanding system: An example of the recognition process. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, Stanford, Cal., 20–23 August 1973.

Rosenberg, S. & Simon, H. A. Modelling semantic memory: Effects of presenting semantic information in different modalities. *Cognitive Psychology*, 1977, *9*, 293–325.

Rumelhart, D. E. & Levin, J. A. A language comprehension system. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975.

Rumelhart, D. E., Lindsay, P. H. & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Schank, R. C. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 1972, *3*, 552–631.

Shepard, R. N. & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1971, *1*, 1–17.

Talmy, L. Semantics and syntax of motion. In J. P. Kimball, (Ed.), *Syntax and semantics* (Vol. 4). New York: Academic Press, 1975.

Winograd, T. Understanding natural language. *Cognitive Psychology*, 1972, *3*, 1–191.