Massachusetts Institute of Technology
Artificial Intelligence Laboratory

AIM 466

April 1978

Modeling Semantic Memory: Effects of Presenting Semantic Information in Different Modalities

by

Steven Rosenberg
M.I.T.
and

Herbert A. Simon
Carnegie-Mellon University

How is semantic information from different modalities integrated and stored? If related ideas are encountered in French and English, or in pictures and sentences, is the result a single representation in memory or two modality-dependent ones? Subjects were presented with items in different modalities, then were asked whether or not subsequently presented items were identical with the former ones. Subjects frequently accepted translations and items semanticly consistent with those presented earlier as identical, although not as often as they accepted items actually seen previously. The same pattern of results was found when the items were French and English sentences, and when they were pictures and sentences. The results can be explained by the hypothesis that subjects integrate information across modalities into a single underlying semantic representation. A computer model, embodying this hypothesis, made predictions in close agreement with the data.

COGNITIVE PSYCHOLOGY 9, 293–325 (1977)

Modeling Semantic Memory: Effects of Presenting Semantic Information in Different Modalities

STEVEN ROSENBERG

Massachusetts Institute of Technology

AND

HERBERT A. SIMON

Carnegie-Mellon University

How is semantic information from different modalities integrated and stored? If related ideas are encountered in French and English, or in pictures and sentences, is the result a single representation in memory, or two modality-dependent ones? Subjects were presented with items in different modalities, then were asked whether or not subsequently presented items were identical with the former ones. Subjects frequently accepted translations and items semantically consistent with those presented earlier as identical, although not as often as they accepted items actually seen previously. The same pattern of results was found when the items were French and English sentences, and when they were pictures and sentences. The results can be explained by the hypothesis that subjects integrate information across modalities into a single underlying semantic representation. A computer model, embodying this hypothesis, made predictions in close agreement with the data.

Research on semantic memory has generally taken one of two paths. Some investigators have chosen to construct computer models which deal with specific semantic domains in plausible ways (e.g., a world consisting of toy blocks). Other researchers have chosen to investigate human performance experimentally (e.g., the comprehension of active and passive sentences). This paper seeks to combine the two approaches by investigating questions about semantic memory experimentally and then using

This work was supported at the Department of Psychology, Carnegie-Mellon University by Public Health Services Grant No. MH-07722 from the National Institute of Mental Health and by the Advanced Research Projects Agency of the Office of the Secretary to Defense (F44620-73-C-0074) which is monitored by the Air Force Office of Scientific Research. A major part of this paper is based on Steven Rosenberg's *Modelling semantic memory: Effects of presenting semantic information in different modalities*, unpublished doctoral dissertation, Carnegie-Mellon University, 1974. Reprint requests should be sent to Steven Rosenberg, The Artifical Intelligence Laboratory, 545 Technology Square, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

ISSN 0010-0285

the results to test a computer simulation that functions as a precise theoretical model of memory acquisition. When applied to the experimental paradigm, the model gives results that are in close agreement with the experimental findings.

This paper has two themes. The first is to explore whether or not a group of related ideas produce, when expressed in different forms (in pictures and sentences; in French and English), a single representation in memory. Before we can talk about whether or not people have different representations, we need to know something about the properties of semantic representations. Thus the second theme of this paper concerns the integration and representation of related semantic information.

Enough is known about memory representation to say that it often reorganizes the surface structure of the stimulus. Thus, Kintsch and Monk (1972), using speed of inferences as a measure, concluded that subjects stored the information from both simple and complex versions of a paragraph in the same form. Reorganization is often influenced by prior expectations. Bransford and Johnson (1973) have shown that passages can be made more or less comprehensible, without changing them, merely by altering subjects' expectations concerning the topics of the passages.

Similarly, Wright (1969) has found that ease of answering an active or passive question was determined by whether or not it corresponded to an active or passive prior statement about which the question was asked. Olson and Filby (1972) have found the same result in a picture verification task.

Although a person's experience with meaningful material will influence his subsequent coding of semantic relations, he sometimes chooses these relations from an underlying system that is relatively parsimonious when contrasted with the surface structure expressions (Chase & Clark, 1972; Clark, Carpenter, & Just, 1973). Some linguists postulate that people use a small set of semantic relations to code all meaning (i.e., case grammars [Fillmore, 1968]).

We shall use conceptual dependency networks developed by Schank (1972) as a simple conceptual base for constructing semantic networks. We postulate that this conceptual base is relatively independent of the form of the material, as long as the meaning is not changed. Thus information expressed either as a picture or as a sentence would be encoded into the same set of relations as long as picture and sentence had the same meaning. Surface structure differences that do not affect meaning would not change the conceptual encoding.

We can rephrase this postulate: Changes in surface structure are important only when they produce changes in semantic structure. This has several testable consequences. For instance, we present a set of related sentences and pictures for learning. Later we express some of the sentences as pictures and some of the pictures as sentences. Subjects should falsely recognize these translations as familiar if the change in surface structure from picture to sentence (and vice versa) has not changed the deep structure conceptual relations.

This same argument holds, of course, for changes in the organization of material within a single modality. If in assimilating new information, we merge related sentences into a single internal representation, then it should be difficult to distinguish new sentences compounded from elements of the underlying representation from old, previously seen sentences. This was the major finding of Bransford and Franks (1971).

We shall use this technique of testing whether or not differences in surface form produce differences in the internal representation to explore our main themes experimentally.

MEMORY FOR PICTURES AND SENTENCES

This experiment explores how people represent information that is expressed in different modes. Specifically, how do people represent pictures as compared to sentences? Pictures certainly seem different from sentences. They are ambiguous in ways that sentences aren't; they often do not explicitly indicate the theme, for instance. At the same time, they present directly spatial relations that often have to be inferred in a sentence.

Are there visual properties inherent in a representation of visually presented information that a linguistic representation does not have? Can we distinguish two separate systems, one visual and one linguistic, for representing and manipulating information? Some studies suggest an affirmative answer: Cooper and Shepard (1973) have presented evidence for the existence of a visual analog space in a letter-matching task. They found that the speed of matching a letter against a target was proportional to the angle of rotation. Paivio (1971) has argued for a dual coding hypothesis in which there are both linguistic and visual systems of representation.

Nevertheless, we will postulate that a group of related communications of the sort used in this experiment is represented in a common conceptual base, regardless of whether they take the form of pictures or English sentences. Failure of the postulate will be revealed in the data by the ability of subjects to discriminate between a sentence or picture stimulus and its "translation" into the other modality.

Failure to achieve accurate translations in the stimulus material (either in this experiment or in the French–English experiment reported in the next section) can cause an apparent disconfirmation of the hypothesis of common semantic representation but cannot cause that hypothesis to be confirmed spuriously. To the extent that subjects are able to discriminate between meanings of a sentence or picture and its translation, they are less likely to recognize one as identical with the other. Hence the design of the experiment is conservative: The data that show inability of subjects to discriminate between sentences and pictures they have actually seen and translations of those items will always *underestimate* the commonality of the semantic representations of those sentences with the representations of *accurate* translations.

To test the hypothesis of common semantic representation, we must control for other important factors besides adequacy of translation. We need an experimental situation where subjects will feel they could have seen a translation. For instance, if all English sentences involve descriptions of animals, it is unlikely that subjects will be fooled by a French sentence about astrophysics. More subtly, subjects may store incidental information with the sentence meanings that allows correct identification. For instance, if an experiment is conducted entirely in French, subjects are unlikely to think that they saw an English sentence. They will make this judgment not because they feel the meaning is wrong but because they know they have not seen anything in English. We need to eliminate the effect of these incidental cues and create a situation where subjects are making judgments on the basis of sentence meaning alone.

A situation where subjects are presented with sentences and pictures composed of a small set of basic elements provides this. When the basic elements are presented in both forms, subjects are forced to decide if the unique meaning of a particular item is the same as the meaning of a previously presented item. Thus, this procedure forces subjects to make judgments on semantic grounds and reduces the effect of incidental or contextual learning.

The paradigm evolved from Experiment II of Bransford and Franks (1971), replicated by Singer and Rosenberg (1973). Bransford and Franks developed idea sets of 12 sentences each by using combinations of four simple propositions. Subjects were first presented with half (6) the sentences of an idea set, balanced for complexity. Then, in a recognition test they were given the other six sentences. It was found that subjects would falsely recognize sentences they had never seen before when these sentences were composed of the same propositions as the original group of sentences. Subjects were as confident in these false recognitions as in correct recognitions. By also presenting some items of an idea set as English sentences, and others as pictures, we were able to test for (false) recognitions of translated items. This device allowed us to explore the representation of related information which is partly expressed in one form, partly in another.

The Experiment

Materials

Four idea sets from Bransford and Franks' Experiment II (1971), previously used by Singer and Rosenberg (1973), were chosen. These all involved concrete ideas that were easily expressed as pictures. A picture was constructed for each of the 12 sentences of each idea set. The pictures were all simple line drawings in black and white. Some minor changes were made in the sentences so that they could be more easily represented as pictures. For example. "large" in "the large window" was changed to "latticed". Figure 1 presents the picture for the four-idea sentence from the idea set modeled in the computer simulation: "The scared cat running from the barking dog jumped on the table."

The pictures were derived from the corresponding sentences using the following guidelines:

(1) The only objects shown in the picture were those representing the nouns mentioned in the sentence.

(2) Objects were drawn as simply as possible, in black and white, on a plain background. An object varied from picture to picture only if the corresponding sentences varied in an attribute (e.g., "the cat" vs "the scared cat"). An identical change was made in an object each time the same attribute occurred in the corresponding sentence.

(3) All implicit spatial relations among objects were kept constant.

(4) The only actions occurring in a picture were those described in the sentence.

There are several issues concerning the equivalence of translations of sentences into pictures whose discussion we will postpone until we have introduced a more detailed model of the representation.

At this point it is convenient to introduce some nomenclature. Items in the recognition list that occurred previously in the acquisition list are called *Literals*. These are sentences and pictures which subjects have seen previously. Items in the recognition list that did not occur in the acquisition list are called *Recognitions*. These are sentences and pictures never before seen but composed out of the same basic ideas as previous items. Lastly, translations of items that occur in the acquisition list are called *Translations*.

All idea sets were presented partly as pictures and partly as sentences in the acquisition phase. A large subject pool was available. allowing the use of a modified Latin Square design. In this design each sentence and picture occurs once in four acquisition lists. Similarly. each sentence and picture is used in one of four recognition lists. Each recognition list is paired with three different acquisition lists. Each sentence and picture of the list functions once as a Literal. once as a Recognition, and once as a Translation in the three pairings.

The acquisition lists were constructed by choosing a one-idea, two-idea, and three-idea sentence from each idea set for each of the four lists. Since there are only three three-idea sentences. one was chosen randomly from each idea set for the fourth list. The pictures corresponding to these sentences were then paired with the sentences so that a sentence and its equivalent picture did not both occur together in the same list. We constructed in this fashion four acquisition lists totaling 24 items: 12 sentences and 12 pictures. In assigning the sentences and pictures care was taken so that in each acquisition list, for each three sentences and three pictures from an idea-set. all four propositions from that idea set occurred with as nearly equal frequency as possible.

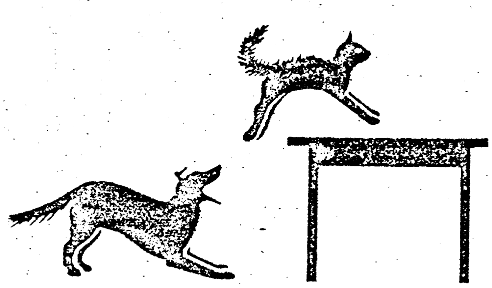


FIG. 1. "The scared cat running from the barking dog jumped on the table."

To form the four recognition lists, we chose sentences and pictures exhaustively from the acquisition lists. To set up a particular recognition list, for each idea set we chose a one-idea sentence from one acquisition list, a two-idea sentence from another, and a three-idea sentence from a third. The pictures corresponding to these sentences were then added to the respective lists, forming four lists of 24 items each. Each acquisition and recognition list was balanced for complexity within pictures and sentences from each idea set. The frequency of occurrence of the four basic propositions of each idea set was made as equal as possible within each mode of each idea set.

Since three of the four recognition lists can be paired with each acquisition list, there are 12 cells in the design. The order of sentences and pictures in the lists was randomized with the constraint that no more than three pictures or sentences or three items of a given complexity could appear in a row. Each sentence and picture occurred in three cells and functioned once as a Literal item, once as a Translation, and once as a Recognition. In each recognition list, there were six Literals (three sentences and three pictures), six Recognition items (three sentences and three pictures), and six Translations (three sentences and three pictures).

Subjects

Four subjects were used in each cell, for a total of 48 subjects. All were fulfilling a course requirement in introductory psychology.

Procedure

In the acquisition phase, subjects were instructed to learn the pictures and sentences as well as possible. The word "learn" was used rather than "memorize" to induce subjects to pay attention to the meaning as well as the surface structure of the pictures and sentences.

Subjects were given 5 sec to view each item of the acquisition set in the T-scope. After viewing the acquisition deck, subjects took the Stroop color card test. This was to provide a brief intervening task before the testing phase. They then saw the items of one of the recognition lists in the tachistoscope. They were instructed to decide as quickly as possible if the item shown was exactly the same as one of the acquisition list items and to press a button indicating their response, "Yes" or "No". Subjects were then told to indicate their confidence in their response on a seven-point scale.

Results

The accuracy, latency, and confidence data were each subjected to an analysis of variance. The independent variables were: Literals versus Recognitions versus Translations, Pictures versus Sentences, and Acceptances versus Rejections (except in the analysis of the accuracy data, where this variable does not exist).

Proportions accepted. There is a significant difference in the proportions accepted for Literals, Recognitions, and Translations. Figure 2 shows these results, $F(2,282) = 90.1$; $p < .01$.

The linear ordering of this main effect has Literals (79.7%) accepted more frequently than Translations (34.9%), which are accepted more frequently than Recognitions (28.7%). Although subjects do not accept as many Translations as Literals, they do accept more than Recognitions,

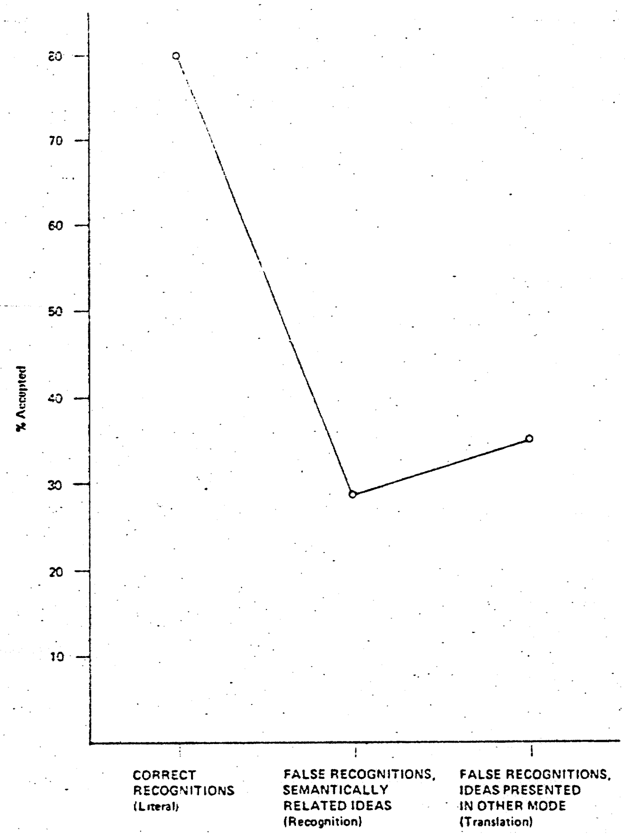


FIG. 2. Picture–Sentence Experiment: Percentages of pictures and sentences accepted for the Literals, Recognitions, and Translations groups.

$t(1,192) = 1.74$; $p < .05$ (one-tailed). Subjects also accept significantly more pictures (53.5%) than sentences (42.2%): $F(1,252) = 11.2$; $p < .01$.

The data were analyzed to determine the effect of sentence complexity. The factors in this analysis were: type of item (Literals vs Recognitions vs Translations); Pictures versus Sentences; and Complexity. Subjects accepted significantly more items as complexity increased, $F(2,270) = 13.3$; $p < .01$: One-idea items, 37%; two-idea items, 53%; three-idea items, 55%. There were no significant interactions with the other factors. Thus, the type of item and whether it was a picture or sentence did not significantly affect how subjects treated increasing complexity of an item.

Latencies. There is a picture (2.24 sec) versus sentence (2.99 sec) difference, $F(1,425) = 62.3$; $p < .01$. Sentences take .75 sec longer to process than pictures. This may be due to reading time. There is no Literal versus Recognition versus Translation difference in processing speed.

Confidences. There is a difference between Pictures (6.1) and Sentences (5.4), $F(1,1128) = 24$; $p < .01$, in the confidence of judgments. This parallels the finding that people accept more pictures and do so more quickly than sentences.

Discussion

The results support the view that pictures and sentences have very similar semantic systems, perhaps the same semantic system, underlying them. Although subjects did not make as many false recognitions of Translations as of Literals, they did accept more Translations than Recognitions, and on average more than one-third of the Translations. The more detailed model, to be presented later, will show that the actual percentages of the three classes of sentences recognized are very close to those theory predicts. The orderings for acceptances of Literals, Recognitions, and Translations were the same in both this and the next experiment, for equivalent conditions.

A bias was found in favor of accepting pictures. Subjects accepted more pictures than sentences (both false and correct acceptances), did so more quickly, and were more confident in their acceptances. This preference for pictures occurs for all categories of pictures. Thus, more Literal, Recognition, and Translation pictures are accepted than are corresponding Literal, Recognition, and Translation sentences. Subjects are more willing to accept pictures than sentences regardless of whether they have seen them previously or not, regardless of whether they are translations or not, and regardless of complexity. There are no asymmetries in the data to indicate a preferred mode of storage.

Bransford and Franks' findings were replicated: Subjects accepted more items, the more complex the sentence or picture. However, consistent with previous research (Franks & Bransford, 1974) subjects did not accept as many Recognitions as Literals. The type of item (Literal, Recognition, or Translation) or its form (picture or sentence) did not interact with this result. Thus subjects give evidence of having similar systems for both pictures and sentences. They integrated a nonconsecutive list of related sentences and pictures in such a way that they were often confused as to whether they had seen a picture or a sentence which expressed the same meaning.

SEMANTIC SYSTEMS OF BILINGUALS

Do people who speak two languages have a single semantic system, or do separate systems underlie each language?

If bilinguals have only a single semantic system, then the language of a sentence is part of the "surface structure." If such a bilingual is presented with a sentence in one language, he should later make a false recognition of that sentence when it is translated into the other language. On the other hand, if bilinguals have two semantic systems, then the same sentence and its translation should result in two separate internal representations and an absence of false recognitions. There is also evidence which supports a mixture of these two positions (e.g., Swain, Note 2). Lambert and Rawlings (1969) argue that certain types of bilinguals, "coordinate" bilinguals, may have separate semantic systems for different parts of each language. This argument is derived from the fact that "coordinate" bilinguals learn each language in a different cultural context. Thus they may grow up in one country and learn to speak in that context. By going to a university in a different country, they may learn a technical vocabulary together with their more advanced understanding, in a different lanugage. Seemingly identical expressions could then have different connotations in each language.

To express the same meaning in two languages involves finding adequate translational equivalents. This is a difficult problem. How can we be sure that the meaning expressed in one sentence has not been changed when we translate this sentence? The answer we have chosen is partly an experimental one and partly one of proper controls. By choosing simple, concrete ideas we increase the likelihood of extensive overlap if not complete identity in meaning. The rest of the answer to this question is rather circular (but not viciously so); if there is only a single semantic system, then we should be able to find good translations that change the meaning only slightly. If there are two semantic systems, then the properties of a representation in one system will always be different from those of a representation in the other system, no matter how good the translational equivalents seem to be. Thus, as in the first experiment, the design is conservative: Poor translations can only work against the hypothesis we are testing.

We can explore this question by using the same paradigm as in the last experiment but substituting French sentences for pictures. The questions asked are these: Do we find the same pattern of results in French as in English? Do native French speakers treat the material in the same fashion as native English speakers? To the extent that this occurs we have evidence of similar semantic systems in French and English in bilinguals. Are the semantic systems identical or only similar? To answer this question, we need to examine how people treat the translated sentences in the recognition phase. Once again, the design seeks to ensure that the reason for a false recognition is semantic similarity, and for a correct rejection, a change in meaning.

Materials

The same four idea sets were used as in the previous experiment. A bilingual French national, who had received his early schooling in England, translated these sentences into French.

A group of bilinguals of French and American origin then inspected the translations, making minor changes. Lastly, pilot subjects were run and questioned about idiosyncracies in the material. Their feedback resulted in a few changes.

Like Bransford and Franks we used sentences composed from a group of four propositions to construct an idea set. Sentences vary in complexity from one to four, in terms of the number of propositions they contain. Of the 48 sentences (12 in each of four idea sets), 24, or half, were chosen for an acquisition list. Six sentences were used from each of the four idea sets: two one-idea, two two-idea, and two three-idea sentences. Each of the four basic ideas of an idea set occurred three times in the sentences chosen for the acquisition list.

The six sentences from one idea set were used only in French and the six sentences from another idea set only in English. Half the sentences from each of the remaining two idea sets occurred in French, half in English. The sentences were balanced for complexity within each idea set and within each language where an idea set occurred in both French and English. The occurrence of the four basic propositions in each idea set was made as equal as possible within each language in which an idea set occurred.

For the recognition phase, a pool of sentences was constructed by forming the following subgroups. For each of the two idea sets that occurred in one language only in the acquisition list, three sentences from the acquisition list together with the translations of the remaining three sentences were chosen. The six sentences of the idea set that were not in the acquisition list, together with their translations, provided a pool of 12 sentences that had never been used in the acquisition list. From this pool, four sentences were chosen in each language.

For each of the two idea sets that occurred in two languages in the acquisition list, three sentences (in one language) from the acquisition list and the translations of all six sentences were chosen. Thus, there were three Translations for each language. Lastly, four sentences never previously seen were chosen from the pool of six remaining unused sentences in each of these two idea sets. In forming these subgroups, the occurrence from an idea set of sentences of different complexity and of the four basic propositions was made as equal as possible.

The pool of recognition list sentences developed with the above procedure was divided into two lists of 27 sentences each. This length was found, through previous replications (Singer & Rosenberg, 1973) and pilot subjects, to be about the longest list possible before subjects start confusing acquisition list sentences with sentences occurring earlier in the recognition list. This list length, together with the available number of bilingual subjects, effectively limited the design to that just outlined.

The two recognition lists were formed by assigning randomly chosen members of each of the subgroups alternately to the two lists until the cell was exhausted. The constraints on this assignment procedure were that frequency of occurrence for each language, and for sentence complexity and occurrence of the four propositions in each idea set, be balanced within each list and as far as possible within the subgroups and sentences from each idea set in a given list. An added constraint was that a sentence and its translation not occur together in the same list. Lastly, the sentence order in the acquisition list and the two recognition lists was randomized. Each recognition set contained six Literals, nine Translations, and twelve Recognition sentences.

Subjects

Subjects were 13 French nationals, and 13 American bilinguals. All were volunteers. The French nationals were upper-level graduate students at Carnegie-Mellon University. The Americans were upper-level graduates majoring in French who had spent at least a semester studying in France.

All were recommended by their instructor as being bilingual. Subjects in each group were alternately assigned to the two recognition list conditions. In testing subjects, those who had difficulty with the material or in understanding instructions were dismissed.

Procedure

The instructions were given in English. After receiving instructions subjects were given the Stroop color card test as a measure of their bilingualism. The mean latencies for the Stroop test were, for the English card: 1 min 65 sec; for the French card: 1 min 70 sec. This difference was not significant, indicating that as a group the subjects had approximately equal facility in French and English.

Next, subjects heard the acquisition sentences, one at a time, from a tape that had been prepared by a bilingual assistant. After hearing each sentence they read five colors from a card, in the same language as the sentence. They then heard a simple elliptical question (e.g., "Who did?" or "Did what?") about the previously heard sentence, still in the same language as the sentence. They wrote down the answer to this question. This procedure was followed to ensure that subjects paid attention to the meaning of each sentence, and the question-answering task served as the ostensible purpose of the experiment.

After proceeding in this manner through the entire acquisition list, subjects were told that there was an additional task to be performed. They then saw the sentences of one of the recognition lists in a tachistoscope. Subjects were given the same instructions as in the previous experiment, except that a five-point scale (as in Bransford & Franks' Experiment II) was used to record confidences. Subjects were discarded if, in the acquisition phase, they repeatedly (more than once in a session) asked to hear a sentence again (they were encouraged to ask for repeats if they did not understand something) or could not answer the elliptical questions correctly. Four subjects were discarded for this reason.

After completing the experiment, subjects were questioned as to whether or not they could recall anything about the material that aided them. This was to test if they had realized that certain groups of sentences occurred in only one language in the acquisition list. Only a few subjects made any comment in response to this question. Every subject who said he knew in which language an idea set had been presented originally was wrong more often than he was right.

Results

Accuracy, latency, and confidence data were each subjected to an analysis of variance. The factors were: Literals versus Recognitions versus Translations, English sentences versus French sentences; Sentences from idea sets that had been presented in two languages in the acquisition list versus Sentences from idea sets that had been presented in one language, Acceptance versus Rejections (not a factor in the Proportions-Accepted ANOVA), and Responses by native English speakers versus Responses by native French speakers. (The one-language case, other-language Recognition sentence responses were discarded.)

Proportions accepted. There is a significant difference among the proportions accepted for Literals, Translations, and Recognitions. Sentences previously seen are accepted most frequently (70.1%). New sentences (falsely recognized) are next (51.6%). Translations are accepted least often (32.9%). The differences among these three measures is significant; $F(2,60) = 17.05$; $p < .01$.

This result must be further analyzed, separating the cases where the idea set had been presented in two languages from those where it occurred in only one. Figure 3 shows the interaction between number of languages and type of recognition sentence. This interaction is also significant: $F(2,60) = 10.8$; $p < .01$.

When the sentences of an idea set had been presented in only one language in the acquisition phase, people make few false acceptances of translated sentences (15.6%), although they accept many Literals and Recognitions (74.2 and 62.7%). However, when the sentences of the idea set had originally been presented in two languages, false acceptances of Translations rise dramatically (to 50.3%). Subjects now accept more translated sentences than Recognition sentences, although this comparison just misses significance, $t(1,34) = 1.3$; $p < .10$ (One-tailed). Acceptances of Literals and Recognitions decrease in the two-language case (to 66 and 40.5%). Thus, in the condition where subjects could have seen the basic elements in both languages (and hence could have seen a particular sentence in either language), they falsely accept many Translations. When they do not receive sentences in both languages, and consequently may not feel that they could have seen a translated sentence, the find it easy to reject them. This result indicates people may use the language of presentation of a set of related (in meaning) sentences as a marker.

Subjects were slightly less likely to accept sentences in their native language than sentences in their second tongue; $F(1,60) < 5.01$; $p = .05$.

The design did not permit a detailed analysis of effects of sentence complexity within the various groups. Although complexity was balanced within these groups, there were often only two sentences of a given complexity (as in the case of a one-idea Literal English sentence, for example). However, an analysis of variance was done to determine if Bransford and Franks', and Singer and Rosenberg's results were replicated, and to see if there were any factors affecting this result. The ANOVA had four dimensions: Literals versus Recognitions versus Translations, native English speakers versus native French speakers, English sentences versus French sentences, and sentence Complexity. The proportion accepted increased significantly with increasing complexity, $F(2,35) = 3.3$; $p < .05$: one-idea sentences, 38.8%; two-idea sentences, 55.1%; three-idea sentences, 55.2%; four-idea sentences, 57.7%. None of the other factors interacted significantly with complexity.

Latencies. Speeds of response show the same pattern of significant differences as acceptances. There is a significant main effect for type of sentence, $F(2,476) = 17.05$; $p < .01$. The fastest responses are made to previously seen sentences (Literals) followed by Translations. Recognitions take the longest time to process.

Confidences. There is a Literal versus Recognition versus Translation difference, with subjects more confident (4.22) of their classification of previously seen sentences than of either Recognition sentences (3.92) or Translations (3.97); $F(2,444) = 3.16$; $p < .05$. There is also a significant difference in confidence between sentences from idea sets originally presented in one language (4.27) versus those coming from idea sets presented in two languages (3.84); $F(2,499) = 7.9$; $p < .01$. Subjects are less sure of their choices when the idea sets originally occurred in two languages.

Discussion

The results show that bilingual subjects have very similar, if not identical, semantic systems for French and English. Subjects integrated the information from a nonconsecutive list of related sentences across languages. As a consequence of this they often accepted translations of sentences seen originally in the other language. Subjects gave the same pattern of responses to test items regardless of the language or type of test sentence, or the native language of the speaker. Thus Bransford and Franks' results were replicated and were not significantly affected by the native language of the speaker or the language of the sentence. Subjects recognized sentences never before seen and were more likely to accept the sentence the more complex it was.

Subjects also made fewer false than correct recognitions, a finding consistent with that of Bransford and Franks (see Franks & Bransford, 1974) and Reitman and Bower (1973). There were no significant interactions between type of sentence (i.e., whether it has been seen before, is a translation, or is semantically consistent with previous sentences) and higher acceptance of more complex sentences. Similarly, the language of a sentence or the native tongue of a subject did not affect the pattern of findings. These results indicate that very similar processes are occurring in the recognition procedures for these groups.

Thus the results support the view that bilinguals have very similar semantic systems for French and English. They are also consistent with the contention that the semantic systems are not only similar, they are identical. In the two-language case, subjects must decide on the basis of meaning whether or not they have seen the sentence previously. If a single semantic structure underlies both languages, the semantic representation of the sentence should be the same as the representation of its translation. In this case subjects should accept the translations as having been seen previously. If a different semantic system underlies French than underlies English, hearing the sentence in one language will not give the same internal representation as hearing it in the other language. The evidence is favorable to the former alternative.

Can our results be explained as an artifact of subjects' guessing strategies? Suppose that subjects stored only about one-half of the stimulus sentences and, during the test phase of the experiment, guessed (with a 50–50 chance of "yes" or "no") at all sentences they did not actually remember. Then the gross data would be rather similar to those of Figure 1, except for the difference between Recognitions and Translations.

This alternative interpretation is implausible for several reasons. First, it is inconsistent with the relatively high confidences subjects assigned to their judgments. Second, it is inconsistent with their rejection (Figure 3) of substantially all translations in the one-language case. Third, it is inconsistent with our more detailed process model and simulation of the precise way in which the internal representation is created from the stimulus sentences.

Subjects did accept fewer Translations (in the two-language case) than Literals, but more Translations than Recognitions. We must ask, therefore, whether or not the failure of subjects always to be confused by Translations or Recognitions can be explained without having to postulate separate semantic systems? We might attribute their partial ability to discriminate Translations to inequivalence of original with translated sentences. However, our later more detailed analysis, using the simulation model, will show that most of the difference can be accounted for without postulating inaccuracies in the translations.

Interactions which would indicate different semantic systems were absent from the data. There were no asymmetries that would indicate that a group of subjects was using a French semantic system only or an English system only. Neither Translations nor any other group were treated asymmetrically. Breaking the data down by native language shows the same pattern of behavior for both groups, although subjects were somewhat faster in their native tongue.

MEMORY EFFECTS: A THIRD EXPERIMENT

As time passes, people often seem to reorganize prose passages (Bartlett, 1967). If this is done, then the distinction between what has been previously seen and what is merely semantically consistent with it, should become progressively blurred. Thus it should be possible, by waiting for an interval, to demonstrate that for the present material, Literals and Recognitions are accepted with the same frequency. Of course, a good model of the semantic information processing in these experiments should be able to explain why we get a particular proportion of acceptances for different types of items (Literals, Recognitions, Translations) in each of the experiments. In the section following this experiment we will introduce such a model, and show how both the integration of related meaningful material across modalities and changes in recognition of this material over time are a function of certain underlying semantic features.

Materials

A set of 24 acquisition sentences was chosen and balanced in the same manner as in the previous experiment. A set of 24 recognition sentences was balanced in the same way as the acquisition group. These were equally divided between Literals and Recognitions in each category.

Subjects

Subjects were 64 experimental psychology students.

Procedure

At the beginning of a 1.5-hr class period, subjects were read the acquisition list sentences one at a time. After each sentence they were asked an elliptical question about it. They wrote down the answers to these on a sheet of paper. This procedure took about 10–15 min. Subjects were led to believe that answering the questions was the purpose of the experiment. The teacher then taught for an hour. The experimenter then returned to the class and administered the recognition list. The same instructions about recognizing an identical sentence were given as in previous experiments.

Results and Discussion

The proportions accepted correctly (Literals) and falsely (Recognitions) are 46.5 and 34.3%, respectively. These differences are not significant, $F(1.17) = 1.56$. Thus, after an hour interval, subjects treated sentences they had seen before and new semantically consistent sentences in nearly the same fashion. Compared to the picture–sentence experiment, the proportion of Literals accepted has decreased and the proportion of Recognitions increased.

THE MODEL

In this section, we will examine the experimental results in the light of a formal theory of the storage of semantic information. The theory has been implemented as a program for the PDP-10 computer and used to simulate subjects in the experiments.

The theory and simultation will provide a more detailed description of the way in which the internal representation is built up from the series of stimuli and will provide explanation for the differences between rates of acceptances of Literals, Recognitions, and Translations that were reported in previous sections.

We postulate that a single semantic system provides a common conceptual base for the understanding of pictures as well as French and English, and we use a parsing scheme developed by Schank (1972) to describe this common semantic representation.

The parsing scheme is used to reduce sentences and their pictorial equivalents to a common conceptual base. The parsing of a picture is quite

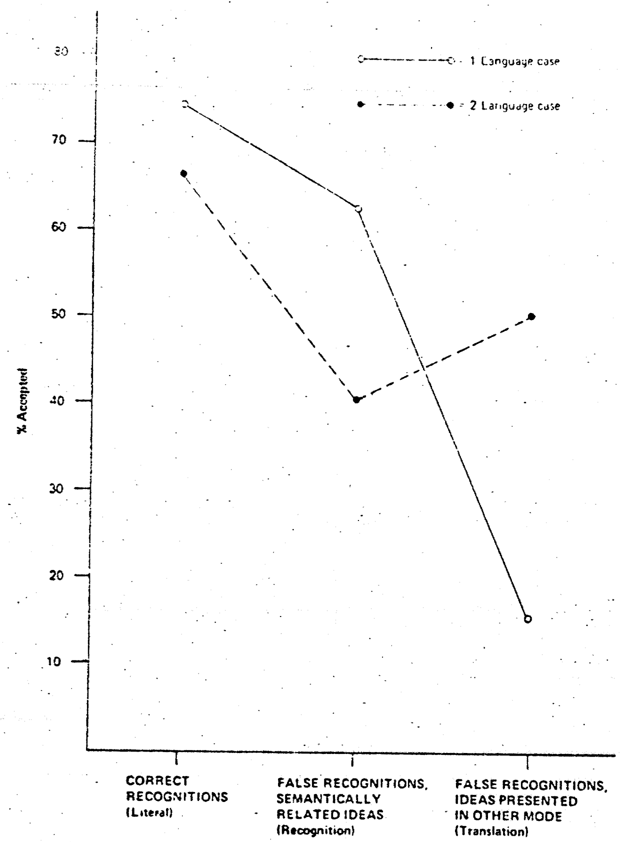


FIG. 3. French–English Experiment: Percentages accepted in the one- and two-language cases for Literal, Recognition, and Translation sentences.

straightforward. The pictures were each created to be equivalent in meaning to a corresponding sentence. (See the design criteria in Experiment 1.) Rather than insist that they were adequate "translations" we argued that only a close correspondence in meaning between a picture and its associated sentence would produce semantic confusion in subjects. The production of such confusion suggests that the pictures and sentences do share a large common core of meaning. Thus, we parse a picture into exactly the same conceptual structure as we would its equivalent sentence. Since we will use these representations in a deterministic model to explain semantic confusions, deviations in the model resulting from actual semantic noncorrespondences should be observable. We will postpone a detailed discussion of this issue until after we have introduced our model.

Although the conceptual base represents both sentences and pictures, the information conveyed by pictures and sentences may differ. We assume an internal lexicon containing our knowledge of lexical entries. An entry defines a *type*, and the meaning associated with the type indicates the possible range of properties this type can have and their possible values. For example, the entry for "cat" will indicate that cats have color and what the possible colors are.

Each actual occurrence of an item belonging to a type is a *token*. A token has connections to the lexical entry of the type to which it belongs, but it also has additional properties. Thus, although cats may be of many colors and sizes, the token representing the cat in a picture will have a specific value for color and size. Upon seeing a cat, the representation of the token for this occurrence of the type "cat" may specify particular values of color and size. On the other hand, the word "cat" creates representation without these values but perhaps specifies certain semantic, linguistic, or typographical properties.

Thus, the information content of representations can be affected by the modality of presentation. This does not imply that there is more than one conceptual base but simply that the information content of a particular occurrence of a type can vary with the mode of presentation. We have expressed this argument in terms of lexical entries that are objects.

Since verbs are also lexical entries capable of a wide range of modification, we would expect the same effect to occur with them. Since, for the present material, there is at least one object that covaries

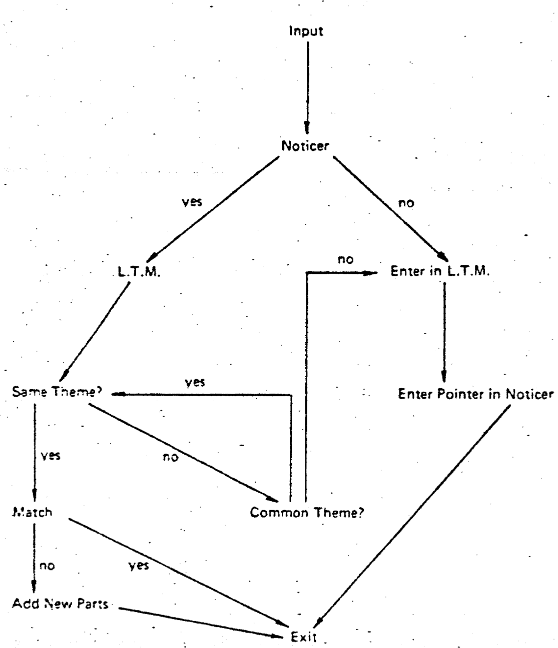


FIG. 4. Acquisition scheme for integrating new sentences and pictures into semantic memory.

with each verb, we have chosen, for convenience, to encode only the tokens for objects with this semantic distinction. Encoding the verbs in the same fashion would produce no observable effect in these experiments.

Why isn't a marker attached to the entire representation or conceptualization rather than to subcomponents? We postulate that the subject integrates related sentences and pictures where possible, rather than keeping multiple representations of common elements in semantic memory. In this semantic memory, a representation will be composed of input from both sentences and pictures. The differences we have postulated are not in the conceptual base representation but in features of the components of this base. Thus the internal representation of a group of related sentences and pictures will contain many tokens with both linguistic and visual attributes, while some, within the same conceptualization, will have exclusively one or the other.

In integrating related material into a semantic structure, the postulated system does not make inferences about connections between conceptualizations. It will only connect two parsings when they share a common conceptualization (if other criteria are also met). This does not imply that people do not make inferences and integrate information on the basis of these inferences. However, when doing so they draw on their knowledge of the content, conversational goals, and so forth. In the sorts of barebones tasks reported in the experiments, we assume subjects are not using such knowledge.

If we program the theory for a computer, by using the same sequences of acquisition and recognition material used in the experiments, the model's behavior can be compared with that of the subjects. The next section will postulate a detailed process for the integration of related ideas upon acquisition and processes for recognizing material by matching it against the internal representation. These processes are used in a computer simulation.

THE COMPUTER SIMULATION

The model was programmed in LISP on a PDP-10 computer. The material used in the simulation was a subset of that used in the Picture–Sentence experiment. One idea set was chosen, "The scared cat running from the barking dog jumped on the table.", and the sentences and pictures in the various conditions from this idea set were input to the simulation. Since 12 different combinations of acquisition and recognition sequences were used in the experiment, the computer simulation had 12 runs, each with a new combination. Special routines were created specifically to create parses of the sentences and pictures, according to the scheme described earlier.

To operate the system, symbols designating the pictures and sentences in a particular acquisition sequence are typed in. Each symbol causes the execution of a set of special routines which create a data structure that is the parse of that item into a conceptual dependency grammar, together with the associated modality-specific information. Each picture is parsed into exactly the same data structure as the corresponding sentence, except for the modality-specific information. The parses are used in the order they are created by the acquisition process.

Acquisition Phase

We will illustrate, step by step, how a sequence of sentences and pictures, in the acquisition phase, causes a semantic representation to develop. (Later, the matching process in the recognition phase will be illustrated.) The examples are taken from an actual sequence used in the Picture–Sentence experiment. Repetitive examples have been dropped.

The decision processes in building internal representations are shown schematically (Fig. 4).

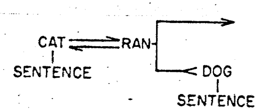
The first part of the system is a "noticer." When the data structure representing the parse for a picture or sentence is examined, components may be recognized, if they have been seen before as topics of acquisition items. A data structure can be entered only through its theme. As a result, the theme of the new sentence or picture will be the first conceptualization examined. On finding this conceptualization, its "picture-producer" (roughly equivalent to the subject) is examined, followed by the "act" (verb) and associated cases (objects, etc.). From here the noticer branches out and investigates conceptualizations connected to the theme conceptualization and then looks at conceptualizations connected to these conceptualiations, and so on. The search is not exhaustive; it stops once any topic is recognized as familiar from previous acquisitions.

The internal representation of each group of integrated sentences and pictures has one of its conceptualizations designated as its theme. The theme is chosen as follows. Each sentence or picture has a theme. (For sentences it is the main clause; for pictures, the theme is the same as the theme of the equivalent sentence.) The theme that occurs most frequently among the merged sentences and pictures in the group is chosen as the theme of the internal representation of the group. This theme may change as additional material is integrated into that internal structure.

If nothing in the new sentence or picture is recognized as a previously encountered topic (the right-hand branch from the noticer in the diagram), the parse of the sentence or picture becomes a new conceptual structure in long-term memory. At the same time, the topic of the structure (defined as the picture-producer of its theme) is added to a list of topics the "noticer" knows about. A pointer is created from this topic to the new internal representation of the sentence or picture. Thus by "noticing" that a component in a new picture or sentence is an already-known topic, the system is able to access a data structure for which that component is the central topic.

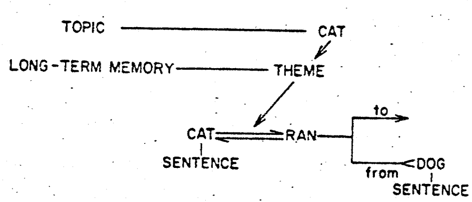
As an example of the process, consider the following sequence of sentences and pictures in the acquisition phase (an actual sequence used in the experiment). The parsing of each is presented; in addition, the semantic representation resulting from this sequence is presented each time this changes.

1. Sentence: The cat ran from the dog.



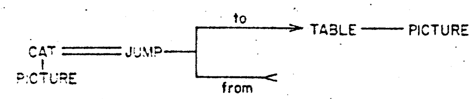
Since this is the first input, no previously occurring topics are noticed as components of the parse. This parsing becomes the first internal representation. There is only one conceptualization in the structure, so the theme of the representation is this conceptualization.

The topic of the theme, "cat," is added to a list of topics (empty till now) the noticer knows about. This topic has a pointer to the theme of the internal representation which is placed in long-term memory (L.T.M.).



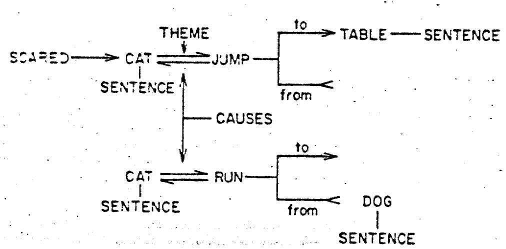
If something is recognized as a topic of a previous input (the left-hand arrow leaving the noticer in Fig. 4), the system will try to merge the new input with the structure in long-term memory that is retrieved through the recognized topic. To do this the system first checks if the theme conceptualization of the new input matches the theme of the internal representation. Each time a new input is merged with an existing structure, a value for the theme of that input is incremented. The theme of the internal representation is the theme that has occurred most frequently in the sentences and pictures that have been integrated to form the internal representation. Thus, as more material is added, the theme of the internal representation can change. (In related material, the theme should reflect the central event being discussed. Since no story is being told or no argument is being made, saliency is judged by the frequency of occurrence of a particular event.)

2. Picture: The cat jumped on the table.



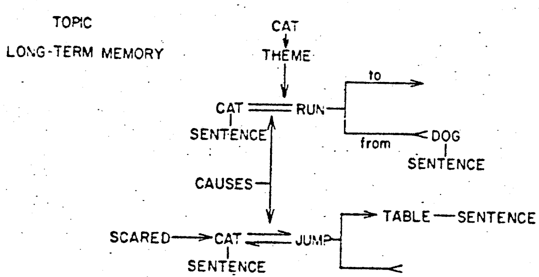
This picture and the previous sentence share the common topic "cat." Following the process on the diagram of the acquisition process, after "cat" is recognized as a previously occurring topic, the conceptual structure this points to is retrieved (the L.T.M. box the left-hand arrow points to). To try to merge the new input into this structure, the system then asks if they have the same theme (same theme box). They do not, as the two conceptualizations are different. The system then checks to see if the theme of either one is a non-theme conceptual component of the other (this is the "common theme" node). This also fails since the two structures share nothing in common besides "cat." The next step in the diagram is to repeat the process done with the first sentence and put this new structure in long-term memory, since it cannot be combined with an existing conceptual structure. However, the topic of this new structure is the same as for the existing structure. Consequently, in this case, the topic will point to two structures. The initial structure is always accessed first. The next structure would be retrieved only if the "same theme" and "common theme" tests failed.

3. Sentence: The scared cat running from the dog jumped on the table.



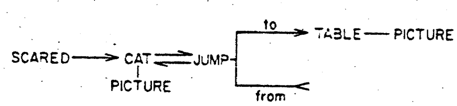
Following the diagram, one again, "cat" would be noticed, and the first internal representation would be retrieved. The two structures do not have the same theme. However, at the next node, "common theme," the system will find that the theme of the internal representation is the second proposition in the new sentence. Consequently, the two representations share a proposition in common, which is also the theme of one of them. The system will now temporarily change the theme of the new sentence to the second proposition. It will then go back to the node "same theme." This time both representations will have the same theme, and this match will succeed. The system then starts to merge the new sentence into the existing structure at the node "match."

This top level description actually covers several processes. In the present case, each proposition in the internal representation is matched against the corresponding proposition in the new sentence. If this corresponding proposition has attributes or cases that are not on the internal representation, they are added to it. In addition, the modality-specific information tags on objects are checked, and if they are new, they are added to the corresponding tag list in the internal representation. Finally the system will notice if there is a proposition in the new sentence that does not exist in the internal representation. This will cause the match to fail and will result in the new proposition being added to the internal representation. After this, the system is ready for a new input. The resulting internal representation will be:

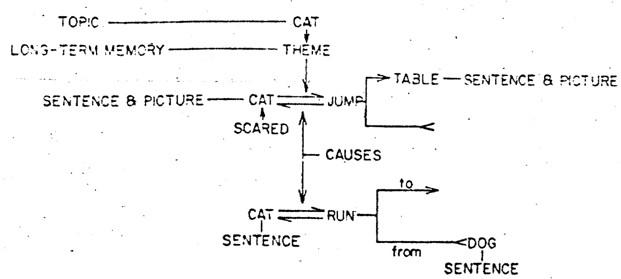


The third sentence had a different theme than the internal representation. However, since this theme has not occurred more frequently than other themes of material that were added to the internal structure (up to this point only one item formed the representation), the theme of the internal structure remains the same.

4. Picture: The scared cat jumped on the table.

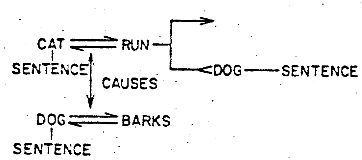


Once again, "cat" is noticed and the same representation retrieved as in sentence No. 3. This is the second time the proposition involving the cat jumping on a table has occurred as the theme of a sentence or picture which can be added to this internal representation. The current theme has occurred only once as the theme of a sentence or picture which has been merged into the internal representation. Thus the theme of the internal representation will change. Otherwise the procedure is the same as that for the previous sentence. In this case the "match" node will succeed, as the only new aspect is the picture tags on the objects. These are added to the corresponding sentence tags, resulting in the following internal structure:

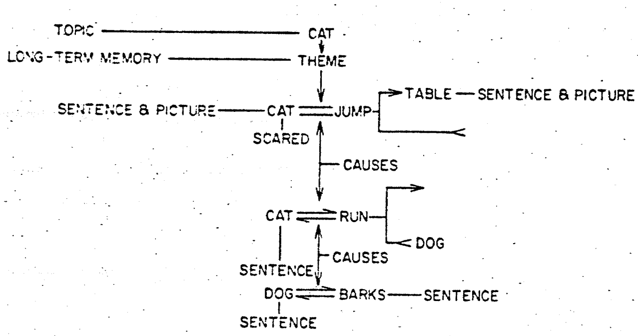


Now the internal representation has some components which are consistent with both modes of presentation, as well as some associated with a particular mode.

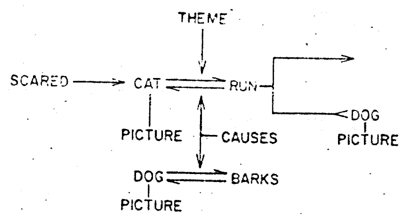
5. Sentence: The cat ran from the barking dog.



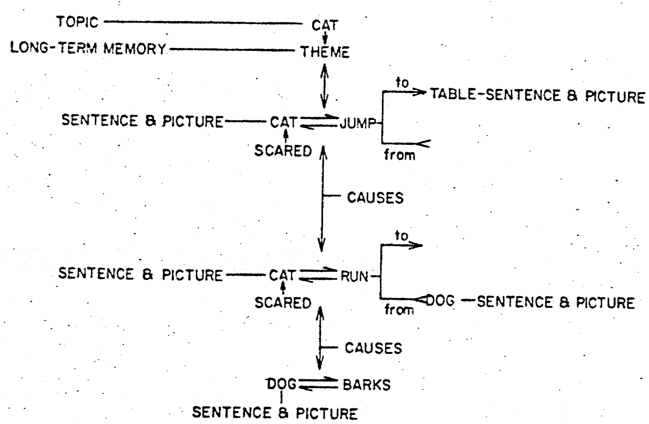
This sentence will be merged into the same internal representation as the previous picture. Now both propositions have occurred equally frequently as themes of material that have been combined with the internal representation. However, the theme of the internal representation stays the same until a new proposition occurs *more* frequently as theme. There is a new proposition in this sentence. Since the sentence shares a common proposition with the internal representation, which is also the theme of both, there is no trouble adding the new proposition.



6. Picture: The scared cat runs from the barking dog.



With the addition of this picture, the theme of the internal representation changes, since "runs" now has occurred more frequently than "jumps." The only other change is that now all objects have tags from both modalities. This is not the case for all sequences. Often the objects associated with at least some of the propositions will have tags for only one modality. Also, in this example, not all sequences create only a single internal structure, as this one does. The final internal representation is:



Recognition Phase

The recognition phase is much simpler than the acquisition phase. Figure 5 illustrates the procedure.

Once again input is provided by typing in the symbol that stands for the intended picture or sentence. This provides the correct parsing to the noticer, which recognizes old topics as in the acquisition phase. If nothing in the input is recognized, the system indicates that it hasn't seen the sentence or picture previously. Occasionally the search will result in a "mistake:" The system will stop searching once noticing occurs but, due to a mismatch, will fail to accept the entire sentence or picture. If noticing then continued, successful recognition might occur. However, since searches start with the most important semantic relations, there will be relatively few false rejections due to this cause.

Suppose that a subject, after receiving in the acquisition phase the sequence of sentences and pictures in our previous example, is presented with

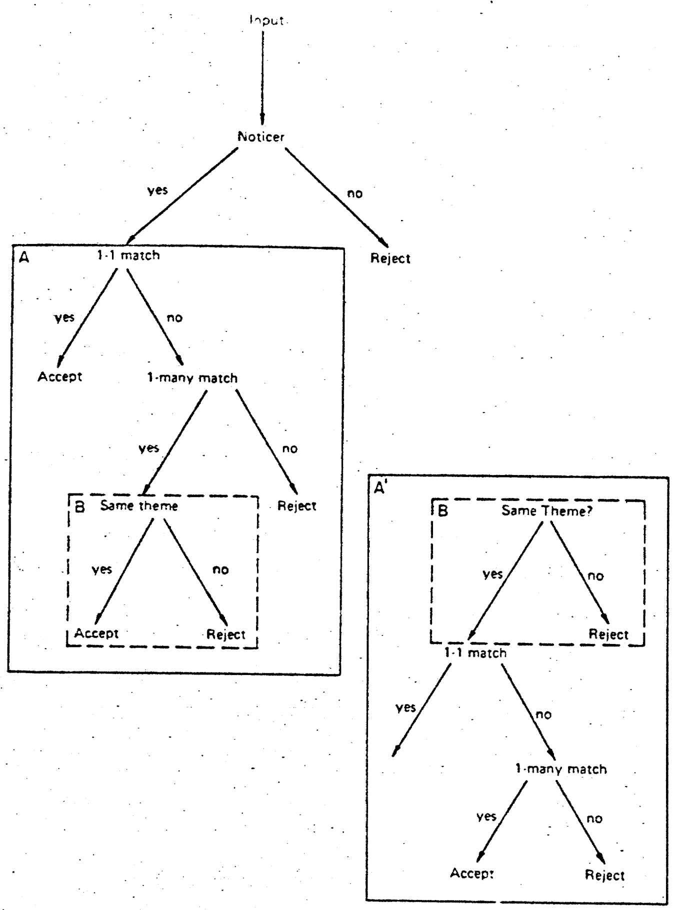


FIG. 5. Recognition scheme for matching a test sentence or picture against semantic memory immediately (A), and after an hour's delay (A').

the picture of "The scared cat running from the dog jumped on the table." This picture is a translation of the sentence presented third in the acquisition sequence. It will have the same parse as this sentence, except that where the sentence has sentence modality tags, it will have picture modality tags attached to objects. Looking at Fig. 5, we can follow the recognition process.

First, the parse is searched, and any familiar topics are noticed. First, the topic of the picture, "cat," will be recognized. Through this topic, the internal representation is retrieved. Each proposition in the new input is matched against the corresponding proposition in the internal structure, (a one-to-one match). Not only must all aspects of the new proposition have a corresponding part in the internal representation, but every part of each of the internal representation's propositions must have a match in the new input. In the present case, starting with the theme proposition, we have a one-to-one correspondence. (The modality tag in a new input need only match one of the modality tags on the corresponding part for the match to succeed.)

Moving on to the second proposition in the new input ("The cat runs from the dog."), we see that it has a matching counterpart in the internal representation. However, the internal representation of this proposition also contains the modifier "scared" attached to "cat." Consequently, the one-to-one mapping fails in this case, and we go to the next node, which involves a one-to-many mapping: That is, for each part of the new input, there must be a corresponding part in the internal structure. However the propositions in the internal structure can have elements, such as "scared," which are absent from the input. This match will succeed.

The simpler match is not enough by itself to cause the picture to be falsely recognized. The picture has been found to be semantically consistent with the internal representation. The internal representation has also been seen before in the picture modality. However, the system also tests if the theme of the new picture is the same as the theme of the internal representation. Here, the theme of the internal representation is the proposition "The cat runs from the dog." The theme of the picture is the proposition "The cat jumps on the table." Since the themes do not match, the translation is rejected.

In a similar way Literals will be rejected if they have a nonmatching theme, and do not have a one-to-one match onto the internal representation. Suppose a translation of the first sentence of our example is presented for recognition. This would be the picture, "The cat ran from the dog." (In the acquisition phase, the sentence was given.) Once again, this picture would have the same parse as the sentence, except that its modality tags would be picture tags. This picture would also fail the one-to-one matching. However, unlike the previous translation, this picture has the right theme, hence would be falsely recognized.

Literals, Translations, and Recognitions are treated in the same way, for the system has no way of knowing if a sentence or picture is a Literal or a Recognition. It happens that fewer Recognitions than Literals have matching themes, or succeed in the one-to-one or one-to-many matches. Sometimes too, a Recognition sequence contains a proposition that is not contained in the internal representation, and this causes rejection. (This can also happen with a Literal if in the acquisition phase it was not combined with the rest of the internal representation.)

Translations are often rejected because the modality tags are not consistent with the stored mode. In our example, if the modality tag attached to "dog" in the proposition "the dog barks" were "sentence" and not "sentence and picture," then any translations of sentences No. 5 and No. 6, above, from the acquisition phase would be rejected. There would be no match for the "picture" tag on "dog" in the proposition "The dog barks," although all other tags would match. Rejection for this reason depends on which sentences and pictures are merged in the acquisition phase and is not due to any inherent separation of picture representations from sentence representations.

Empirical Tests of the Simulation

The most interesting question is whether the simulation model of semantic memory predicts the Literal versus Recognition versus Translation acceptance rates of subjects, using only the principles described previously. The acceptance rates for subjects and simulation, on the material used in the simulation, were, respectively, 78.1 versus 66.6% for Literals, 33.3 versus 33.3% for Recognitions, and 42.7 versus 45.8% for Translations. The simulation provides a close approximation to the human data, within 5 to 10%. The experimentally observed difference in acceptances between Literals and Recognitions is produced, and, more important, the difference between Literals and Translations is also produced. Thus the idea that a single system represents semantic relations, independent of form of presentation, is supported.

How well does the simulation correspond to the other experimental findings? For instance, does it replicate Bransford and Franks' results? The simulation replicates their findings to the same extent that the experiment does: The longer the sentence, the more likely it is to be accepted. This occurs for various reasons. The longer the sentence or picture, the more likely it is to match completely onto an internal structure. It is also more likely to contain the theme of the internal representation.

If our theory is robust, it should be able to model our findings in the time-delay experiment, and in the French–English experiment (which took an hour to perform) in a straightforward way. In fact, it will do so if we make one simple change in the recognition processes to represent the memory reorganization that occurred during the time delay.

We have postulated that the theme of an internal representation has an important function in recognition processes. We further postulate that as time goes by, the theme will become a more important decision criterion. Whereas initially the model first checked for semantic consistency, and then for thematic identity, after a period of time these operations become reversed. Thus, after a wait, one is more likely to ask "Is the new sentence about the same thing as a previous group of items?" before bothering to match the fine semantic structure. This is indicated in Fig. 5 where Box A′ now replaces Box A in the recognition process. The sole difference between A and A′ is that Sub-box B has moved from the bottom to the top, or beginning, of the decision process. By moving this "same theme" node from the bottom to the top of the decision tree, we insure that matching of themes will be an important criterion in all recognitions. Thus, in Box A, an item can be accepted if it is a particularly good match (i.e., one-to-one match) without ever comparing themes. In Box A′ this is no longer true, since the theme is always matched first. In this latter case, if the themes match, either a one-to-one match or a one-to-many match will result in acceptance of the item.

The results with our revised simulation, when applied to the 1-hr delay experiment are: for Literals, 68.3 versus 66.7%; and for Recognitions, 34.5 versus 33.3% for subjects and the simulation, respectively. Once again there is a good fit to the data. Similarly, the results with the simulation, when applied to the French–English experiment (which took about an hour to perform) in the one-language condition, where Translations are almost always rejected (for subjects and simulation, respectively) are: 67.3 verses 66.7% for Literals, 55.5 versus 54% for Recognitions, and 15.4 versus 0% for Translations. Again we find a close fit to the data. The simulation rejects the Translations. No translation will successfully match the semantic representation since no token in the representation will have a modality tag consistent with the translated form of the token.

GENERAL DISCUSSION AND CONCLUSION

The theory developed here for internal representations is a simple one and would have to be expanded to become a general model of semantic memory. For instance, in long-term memory no inferences are made in merging input sequences with internal structures; only the immediate information from the parse is used. The theme of an internal structure is simply the most frequently occurring theme in the input strings that formed that structure. Since the test sentences and pictures do not develop an argument, tell a story, or fit within a larger context, these assumptions are reasonable. However, they would have to be changed for ordinary discourse.

Pictures and sentences are represented in the same fashion. Tokens for objects can have markers indicating the different modalities, but no other distinction is made between visual and linguistic input; the semantic relations are represented in the same way, and picture and sentence meanings are merged into single structures.

The model provides a good quantitative explanation for the results of three experiments. The experiments showed that there were differences between rates of acceptance of Literals and Recognitions, and between Literals and Translations, and they also showed that semantic systems with similar properties underlay the different modalities. The computer simulation of the model was able to produce these results without assuming either separate representations for information in different modalities or retention of copies of particular sentences and pictures.

In evaluating the model we should keep in mind that it is not a model of a specific subject but of an "average" or "ideal" subject. Not all subjects use the same processes or respond in the same fashion to each particular test item. Consequently, we can expect the model to agree with subjects only to the extent that they agree with each other. For instance, if a subject accepts a particular sentence or picture from the idea set used to test the model, what proportion of the other subjects also accept that item? For Literals, Translations, and Recognitions, these proportions were 78, 60, and 46%, respectively. (Over all four idea sets the respective proportions are 81, 55, and 47%. Thus the idea-set chosen for the simulation was representative in this respect.) For each item the computer model accepts, the proportion of subjects who agree with that choice are 75, 57, and 42% for the Literals, Translations, and Recognitions.

The fit is equally good if we consider items that the model rejects. In this case, the proportions of subjects agreeing with the model are 16, 69, and 69% versus 35, 72, and 73% for the proportion of subjects agreeing with each other (for Literals, Translations, and Recognitions, respectively). Thus, on both acceptances and rejections, the computer model agrees with subjects concerning individual items almost as frequently as subjects agree with each other.

The issue of fit arises particularly strongly in the translation of sentences into pictures. Since pictures are ambiguous in certain ways, how can we be sure subjects encode them in the fashion we have described? The answer is that probably subjects do not always encode in this fashion. The model only requires that as many subjects agree with it in encoding as agree with each other. The pictures were constructed from the sentences by a set of rules (See Experiment 1) designed to make alternative encodings unlikely. However, let's consider some possible translation errors.

First, suppose the picture contains extra information not in the sentence. We can distinguish two sorts of information, extra visual information about objects and extra spatial information about the relations among objects. The first sort is already incorporated in the model and need not concern us. The second sort will always be present in pictures, never in sentences. Since pictures always followed the same format, this information did not change from one picture to another containing the same objects. Consequently each translation of a sentence into a picture will have that information added, corresponding exactly to information in the other pictures. No sentence will contain this information. In the model, the information would add extra propositions to a semantic structure. These never match against a sentence, since the sentence does not contain this type of information; they always match against a picture, since the information does not change from picture to picture. Consequently, although subjects may or may not encode this pictorial knowledge, it will not affect acceptance or rejection of recognition sentences and pictures.

Two other types of information are ambiguous in pictures: actions and themes. Although only the actions given explicitly in the sentences were used, and an attempt was made to make these as unambiguous as possible, subjects may have used an alternate encoding of pictures. For instance, a possible encoding of Fig. 2 might be "The dog that was chasing the scared cat jumping on the table was barking," or some other variation. Of course, not all such surface-structure variations produce different conceptual representations, but many of them do and also produce different themes.

In considering this problem, we must take account of the context in which encoding occurs. We know that the encoding of a particular item is affected by preceding items, (e.g., Bransford and Johnson, 1973; Wright, 1969; Olson and Filby, 1972). The context will bias a subject to encode pictures in the form postulated. We propose that in fact most subjects do this. However, we are *not* saying that the pictures have a single encoding or that all subjects use this encoding—only that given our experimental conditions most subjects do so. Subjects who do not are likely to disagree with the model and with other subjects over the acceptance or rejection of a particular item. The data show that about a third of the subjects do not agree on any specific item either with the choices of the model or with the choices of other subjects.

Having presented our model, it is appropriate to consider alternative explanations of our results derived from dual coding models It can be argued that Translations of items acquired in one modality are functionally Recognition items in the other modality. Consequently, small differences between Translation and Recognition items are consistent with a dual coding model in which there are separate, noninteracting semantic systems for the visual and verbal modes.

We argue against a dual coding explanation of our results on both experimental and theoretical grounds. First, although the Translation–Recognition difference was not as large as the Translation–Literal difference, it was significant in the Picture–Sentence experiment and just missed significance in the French–English experiment. Rosenberg (Note 1), using different materials also found a significant difference between Recognition and Translation items. Consequently the Translation–Recognition difference was significant in two out of three experiments, and the direction was as predicted by our model in all three.

A more interesting finding, however, follows from a consideration of our model's specific predictions as to how individual items are recognized. For instance, there should be a difference in the way Translations are processed as compared to Recognitions. The processing of Translations should resemble that of Literals rather than Recognitions.

Consider the use of the theme in the Picture–Sentence experiment. In our model, an item having a good (one-to-one) match and consistent modality tags is accepted without themes ever being compared. Items with poorer (one-to-many) matches must, however, also have matching themes in order to be accepted.

In this regard, Literals should generally have good, (one-to-one) matches, since they have been seen previously. Consequently the theme should not be an important factor. If we contrast Literal items containing the theme (82.5%) with those that do not (77.5%), we find similar levels of acceptance and no significant difference, $F(1,279) = .51$.

Translations should behave like Literals in this regard, given our model of a single semantic system. Consequently levels of acceptances for Translations containing the same theme as the internal representation (40.6%) should not differ significantly from those which do not (39%), and this is indeed the case.

Recognition items, being new combinations of elements, should often have a poorer (one-to-many) match. This should make the theme a more important criterion for Recognition items. Those Recognition items having the proper theme (39%) should be accepted more often than those which do not (22.5%). This difference is highly significant, $F(1,279) = 15.2$; $p < .001$.

The difference in the way Translations and Recognitions behave indicates that they are not processed in the same fashion, regardless of overall levels of acceptances. A model based on separate noninteracting semantic systems cannot predict this effect.

Our results also exclude most forms of a dual coding model in which the two systems interact. For instance, if subjects encode items dually during acquisition, both forms of a Literal will be present. Subjects should then accept Translations as frequently as Literals, which they do not. On the other hand, if dual encoding doesn't occur, consider the recognition process. If a subject matches an item against both codes during the recognition process, Translations should always be rejected. This also does not occur. Consequently we are driven back to the dual coding model, already rejected above, in which separate, noninteracting semantic systems exist for each modality. We cannot claim to have considered all kinds of dual encoding models, however elaborate, that might be imagined, but we believe those discussed and eliminated above are the most plausible. Most proponents of multiple-representation models argue for dual encoding on acquisition, as in the interacting model we have rejected. Paivio (1971) for instance argues that linguistic competence for concrete items is initially dependent on a substrate of imagery and that imagery plays an integral role in linguistic comprehension. He presents evidence that both codes are available for sentences and pictures involving "concrete memory representations."

Finally, we would like to offer some comments about our basic paradigm, derived from the work of Bransford and Franks. The Bransford–Franks procedure has been criticized, principally on the grounds that since their result can be produced with nonmeaningful material such as nonsense sentences and letter strings (Katz & Gruenewald, 1974; Reitman & Bower, 1973), it is not the result of semantic integration, but of simple guessing strategies (Katz, Atkeson, & Lee, 1974). We must ask ourselves if semantic memory is different from memory for nonmeaningful material. In the model we have proposed, memory involves an interaction between acquisition and recognition processes, and complex internal representations of meaningful material. Letter strings and nonsense sentences obviously produce much simpler internal representations than those we have considered. The real issue then is: Do these simpler representations interact with the acquisition and recognition processes in a different fashion from more complex material? Work on this question is currently in progress.

In the opposite vein, Bransford and Franks' results have been criticized as artifactual since a change of instructions can change the findings. For instance, Katz (1973) found that asking subjects to recognize whether or not sentences meant exactly the same thing as previously presented sentences (and not whether or not they had actually seen such sentences) reduced the effect of sentence complexity upon confidence. The change in instructions would affect the recognition decision processes of memory, leaving the other components unchanged. And in fact there are intuitively satisfying changes which could be made in our postulated decision processes that would produce such a reduction in effect. For instance, under such instructions, a subject might decide that as long as the information content of a sentence matched an internal representation, the occurrence of a common theme was no longer important.

In order to understand behavior on cognitive tasks we must have precise models of the whole set of processes (such as "semantic integration") involved. Otherwise, we will frequently find ourselves unable to model a change in condition that changes our results. There is then a great temptation to start over again and build a new theory for each new experiment. We have tried to show in this paper that by building our theory into a model that can be used to explain behavior over a range of experimental situations, we provide a means for gradually generalizing our theoretical explanations to wider and wider domains. In constructing the theoretical model we have sought to enhance that generality by using, wherever possible, components (e.g., Schank's parsing scheme) that have already been employed in previous work.

REFERENCES

Bartlett, F. C. *Remembering*. Cambridge, MA: University Press, 1967.

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

Bransford, J. D., & Johnson, M. K. Consideration of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Clark, H. H., Carpenter, P. A., & Just, M. A. On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Cooper, L. A., & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Franks, J. J., & Bransford, J. D. A brief note on linguistic information. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 217–219.

Fillmore, C. J. The case for case. In E. Bach and R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehard & Winston, 1968.

Katz. S. Role of instructions in the abstraction of linguistic ideas. *Journal of Experimental Psychology*, 1973, **98**, 79–84.

Katz, S., Atkeson, B., & Lee, J. The Bransford–Franks linear effect: Integration or artifact? *Memory and Cognition*, 1974, **2**, 709–713.

Katz, S., & Gruenewald, P. The abstraction of linguistic ideas in "meaningless" sentences. *Memory and Cognition*, 1974, **2**, 737–741.

Kintsch, W., & Monk, D. Storage of complex information in memory: Some implications of the speed with which inferences can be made. *Journal of Experimental Psychology*, 1972, **94**, 25–32.

Lambert, W. E., & Rawlings, C. Bilingual processing of mixed-language associative networks. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 604–609.

Olson, D. R., & Filby, N. On the comprehension of active and passive sentences. *Cognitive Psychology* 1972, **3**, 361–381.

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars and concepts. *Cognitive Psychology* 1973, **4**, 194–206.

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

Singer, M., & Rosenberg, S. T. The role of grammatical relations in the abstraction of linguistic ideas. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 273–284.

Wright, P. Transformations and the understanding of sentences. *Language and Speech*, 1969, **12**, 156–166.

REFERENCE NOTES

1. Rosenberg, S. *The representation of concepts in different forms*. (A. I. Memo 382). Cambridge, Massachusetts: M. I. T., Artifical Intelligence Lab, 1976.
2. Swain, M. *Bilingualism as a first language*. Unpublished doctoral dissertation, University of California, Irvine, 1972.

(Accepted November 23, 1976)