

INSTITUTE FOR THE
STUDY OF
INTELLECTUAL
BEHAVIOR

RECOGNITION MEMORY FOR ASPECTS OF DIALOGUE

Elizabeth Bates, Mark Masling and Walter Kintsch

Program on Cognitive Factors in Human Learning and Memory Report No. 69

Institute for the Study of Intellectual Behavior

University of Colorado

Boulder, Colorado 80309

September, 1977

ABSTRACT

Recent research has suggested that memory for surface form in natural discourse may be more robust than laboratory studies of connected prose have led us to believe. This experiment examined recognition memory for anaphoric vs. explicit reference in a 20-minute videotaped drama. The results demonstrate significant memory for meaning, operationalized as the ability to reject a false paraphrase. Furthermore, there was significant memory for surface form for several types of reference, including pronouns vs. proper names, role vs. proper names, and elliptical vs. full clauses. In general, surface memory tends to be higher for explicit reference than for anaphoric utterances. Finally, there were systematic and significant biases toward several surface forms in control groups that were guessing about the relative naturalness of alternative utterances. The surface-meaning distinction, at least in natural discourse, should be re-assessed as a distinction between semantic meaning, and the pragmatic function of various surface forms. Some problems concerning retrieval vs. reconstruction in memory are considered.

This study is part of a series of investigations into the nature of memory for naturally-occurring linguistic materials, in natural settings (see also Kintsch and Bates 1977; Keenan, MacWhinney and Mayhew, in press). The purpose of these studies is to examine whether the processes used in natural linguistic memory are the same as those used in the laboratory, and in particular to investigate memory for those structures that are characteristic of natural discourse.

Recent research on memory for discourse has focused our attention on several processes that were less apparent in traditional research on individual sentences or words. For example, Bransford, Barclay and Franks (1972) have demonstrated that memory for a series of related sentences is strongly influenced by a process of semantic integration. In this process, subjects construct inter-sentential meanings based on both explicitly-presented information, and on information that was implicit in the relations among the stimulus sentences. When asked to determine which sentences actually occurred in the stimulus set, subjects are often unable to distinguish inferences implicit in the text from material that was explicitly presented. The theory of semantic integration has in turn led to two further emphases: (1) the distinction between memory for surface form and memory for meaning, and (2) the distinction between memory-as-retrieval, in which some sort of trace (surface or meaning) is recovered from storage, vs. memory-as-reconstruction, in which meaning or surface material is recognized or recalled not through retrieval of traces from storage, but rather by rebuilding the input on the basis of knowledge about probable structures in a given context.

With regard to the first distinction, it has been argued that memory for surface form in discourse is extremely weak and short-lived. True paraphrases

(i.e. paraphrases consistent with the meaning of the text) are indistinguishable from the actual input sentence within 80 syllables (Sachs 1967) or within four to eight intervening sentences (Garrod and Trabasso 1973). Furthermore, with regard to the second distinction, researchers like James, Thompson and Baldwin (1973) have argued that the apparent recall or recognition of surface forms may be based on reconstructive processes rather than on "true" retrieval. A subject can often use his knowledge about the construction of sentences in general, or in a given context, to choose correctly between competing expressions for a given semantic structure--without having actually seen or heard the target sentence at all. At the very least, this finding suggests that in recognition memory research, we must be very careful to examine the natural biases favoring one paraphrase over another.

In general, the predominance of semantic memory over memory for surface form has been explained in terms of the ecological validity of semantic integration in memory for natural language--assuming that the passages used in these laboratory studies are an approximation to language in natural settings, and that the expectations subjects bring into the laboratory do not radically alter the comprehension and memory processes that they use with these texts. The argument has been made that, if anything, the artificial aspects of the laboratory setting should enhance surface memory beyond its level in natural settings. In a summary of the literature on memory for prose, Clark and Clark (1977) conclude as follows:

Normally, people "study" speech by listening to it for its meaning and by discarding its word for word content quickly. They try to identify referents, draw inferences, get at indirect meaning, and in general build global representations of the situation being described. When they later try to remember this speech, they fail miserably on its verbatim content, and they confuse two names for the same referent, a sentence and its implications, and one piece of a global representation

with another. Yet when they have to, people can "study" speech word for word and later recall it verbatim. Memorization, however, usually requires repetition and special concentration on the surface features of the speech to be remembered.

However, two recent studies have suggested instead that recognition memory for surface form in natural settings is in fact surprisingly robust, without memorization or, for that matter, awareness that memory will be tested at all. Kintsch and Bates (1977) carried out two experiments on recognition memory for statements from a classroom lecture, including topic sentences, detail sentences, and a category of jokes and other statements extraneous to the "text act" of the lecture. In both experiments, the lectures were given during the usual class period, and students were unaware that they would be tested for sentence memory (except, of course, for the usual expectations that classroom material would be relevant for course examinations). In the first experiment, after a 48-hour delay students were able to distinguish target statements from true paraphrases for topics, details and for jokes and other extraneous statements. In the second experiment, after a 5-day delay surface memory was significant for jokes and extraneous material, although it had disappeared for topics and details. In other words, surface memory for discourse in a classroom lecture is much stronger than surface memory in laboratory experiments; furthermore, memory for surface form is a function of the role of a statement in the whole text unit.

In a similar study, Keenan, MacWhinney and Mayhew (in press) tape-recorded a faculty lunchroom conversation, and administered a recognition memory test to the unsuspecting participants 36 hours after the conversation had occurred. After this delay, subjects were able to distinguish target sentences from true paraphrases for utterances high in what Keenan *et al.* term "interactive value", e.g. figures of speech, mock insults and jokes. Descriptive statements low in interactive value

showed significant memory for meaning (i.e. rejection rates for statements that did not occur in the conversation), but no significant memory for surface form. Control experiments demonstrated that the advantage accrued to high interactive utterances is a function of participation in the whole discourse, rather than the relative "memorability" of the individual utterances taken out of context. In a list-learning experiment, with the same utterances presented in random order to non-participants in the conversation, there was no difference in memory for surface form for high vs. low interactive utterances.

In the present study, we have focused on memory for anaphoric processes in dialogue, in particular pronominalization and ellipsis vs. explicit reference. There are two reasons for this choice. First, anaphora is a pervasive process in natural discourse. High-frequency use of pronouns and elliptical reference characterizes natural conversations, and distinguishes them from the more formal types of prose typically used in studies of discourse memory, e.g. stories and paragraphs. Second, the question of memory for surface form seems particularly salient with regard to anaphoric processes. If the function of pronouns and other short forms is merely to identify previously established referents, why should we retain any information other than knowledge of the referent itself? If surface form for anaphora is retained over any significant period of time, we may have to re-assess our definition of the nature and role of surface form in semantic integration of natural discourse.

In order to have some measure of control over these subtle processes in a large enough sample of subjects, we have had to take one step backward into the laboratory. In this experiment, the stimulus is a videotape of an afternoon television drama (a "soap opera" entitled Another World). This piece of discourse was chosen over other candidates because the afternoon serial, in comparison with other dramatic forms, more closely resembles natural conversation in the number of

assumptions that are made about ongoing knowledge of characters and events. Hence this dramatic form mimics the proportion of "old" vs. "new" references to characters and events that typically occur in natural conversation. Obviously, the subjects in the study were not participants in the discourse, and the research was necessarily carried out in a laboratory setting. However, until the recognition memory test was presented, there were no cues to suggest that a memory test was at hand. The videotape was rich enough in explicit vs. anaphoric reference to yield a large enough item pool for testing, while at the same time approximating natural conversation as closely as possible.

METHOD

Subjects: 120 college students participated as subjects in small groups. All obtained credit for an introductory psychology course.

Materials: A half-hour daily segment from an afternoon dramatic serial was videorecorded in March of 1976, six months prior to its use in the present experiment. After editing to remove commercial messages, the entire segment lasted approximately 20 minutes, divided into six separate scenes with changes of setting and characters. The tape was transcribed and checked to insure correct wording.

From this transcript, 43 target sentences were chosen, distributed evenly across the six segments. The utterances were chosen to reflect six sentence types, in three reciprocal sets. The pronoun set contained expressions that referred to main characters by name, and expressions that referred to main characters by pronoun. The clausal set contained utterances expressing information with fully-formed clauses, vs. utterances referring to similar information with elliptical clauses. The role set contained utterances referring to main characters by role (e.g. "his wife") vs. utterances referring to main characters by name, or by pronoun. This last set, though infrequent in the text, provides a particularly interesting contrast in anaphora, since in the pronoun set proper names contain

more explicit information than their reciprocal pronouns, while in the role set proper names are less explicit, marked, or informative than the role references.

For each of the 43 target utterances, a multiple-choice item was constructed containing the target utterance, a true paraphrase, and a false paraphrase. The true paraphrase contained the same information as the target but reversed the direction of the surface form. In other words, a pronoun sentence was paraphrased with identical wording, except that the character's name was substituted by a pronoun. A name sentence in the same set was also paraphrased with identical wording, except that a pronoun was substituted for the character's name. Similarly, in true paraphrases of the clausal set, full clauses were substituted by elliptical clauses with the same meaning. Elliptical clauses were paraphrased with an equivalent full clause. Finally, in true paraphrases of the role set, the character's role was paraphrased with his proper name or a pronoun, while in name sentences in the same set, names were substituted by an appropriate role term. The false paraphrases also maintained exactly the same wording as the targets and true paraphrases, except that a false referent was substituted in the same name, role or clause slot undergoing contrast. For example, a typical multiple-choice item from the pronoun-name set would be the following:

- (a) I wanted to get that Pendleton work done while he was out of the office (target)
- (b) I wanted to get that Pendleton work done while Robert was out of the office (true paraphrase)
- (c) I wanted to get that Penuleton work done while Willis was out of the office (false paraphrase)

Similarly, a typical multiple-choice item from the clausal set would be the following:

- (a) We're doing everything we can to make sure that she does, Aaa
(target)
- (b) We're doing everything that we can to make sure that she keeps the
baby, Ada (false paraphrase)
- (c) We're doing everything that we can to make sure that she regains
consciousness, Ada (true paraphrase)

All of the false paraphrase contained names or information relevant to the script, but either untrue or inappropriate in the particular utterance in question. There were different numbers of items in each item type. In the pronoun set, there were 10 pronoun-target utterances and 10 name-target utterances. In the clausal set, there were 10 elliptical-clause-target utterances and 8 full-clause-target utterances. Finally, in the role set there were only two role-target utterances and 5 name-target utterances. This latter set was necessarily small, reflecting the relatively low frequency of this type of construction in the text.

The multiple-choice items were typed in random order in a recognition-memory test booklet. In addition, a 250-word synopsis of the story, episode by episode, was constructed for use with half the subject. This synopsis was also accompanied by a list of the full cast of characters and their relations to one another.

Procedure: The 120 subjects were assigned randomly to four groups, with 30 subjects in each. Group I received neither the synopsis nor the television program. They were simply asked to judge within each multiple-choice item the utterance most likely to have occurred in an afternoon television serial. Group II did receive the synopsis and cast of characters, but did not view the program. They too were asked, after reading the synopsis, to select the most likely alternative within each of the multiple-choice items. Group III viewed the videotape but did not receive the synopsis. Their instructions were to circle the alternative within each item which they had actually heard in the program. Group IV read the synopsis prior to viewing the tape, and were then given the multiple-choice test with the

same instructions as Group III. This procedure yields a 2x2x6 design, with program/no program and synopsis/no synopsis as between-subject factors, and the six item types as within-subject factors.

Within each of the four groups, subjects were tested in smaller groups of 4 to 6 members. Testing took place in a laboratory room with no visible apparatus other than a large television monitor and a playback machine. The two Program groups (III and IV) were simply told that they would see a television program and be asked about it afterwards. Although no deception was involved, there were also no cues to suggest that the purpose of the experiment was to assess sentence memory.

We should note that the test items are not thematically independent of one another. Merely by reading through the test booklet, without a synopsis, a subject could obtain some idea of the plot and the relations among characters (e.g. that Rachel is in some sort of medical trouble, and that either her husband or her father is worried about her). Hence neither of the control groups was expected to perform at chance levels in rejecting the false paraphrases. However, there is nothing in the randomly-ordered test items to favor one surface form over another, e.g. to distinguish "Rachel" from "her" in a given sentence. If control subjects perform beyond chance levels in distinguishing true paraphrases from targets, their performance presumably reflects systematic biases about the nature of conversation, or at least the conversations that are thought to occur in television serials.

RESULTS

Because the six item types contained unequal numbers of utterances, percentage scores were used rather than absolute frequencies. Memory for meaning was calculated for each subject, for each item type, by subtracting the percent choice of false paraphrases from 100%. Hence if performance were random, memory for meaning would average 66%. Memory for surface form was calculated for each subject, for each item type, by subtracting the percent choice of the true paraphrase from the percent choice of the target utterance. In this case, random

performance should yield surface memory scores of around 0%. Two separate analyses of variance were carried out on arcsin transformations of the percentage scores, one analysis on memory for meaning (the likelihood of choosing either the target utterance or its paraphrase), the other on memory for surface form (the difference between the probability of choosing the target minus the probability of choosing its paraphrase).

Memory for Meaning: Analysis across Subjects

Table 1 presents the memory for meaning scores for all four groups, for each of the six item types. The analysis of variance on these scores yielded significant main effects for all three factors.

First, there were two significant between-subjects effects, $MS_{\text{error}} = 351.5$. The main effect of having seen the program was significant beyond $p < .001$, $F(1,116) = 98.37$. Hence subjects who have actually seen the videotape are better at rejecting false paraphrases than those who have not. However, the two control groups were also performing at well beyond the 66% chance level (see Table 1). There was apparently sufficient information available in the internal structure of the test booklet alone to facilitate rejection of false alternatives. The synopsis also had a significant effect on memory for meaning, at a level beyond $p < .025$, $F(1,116) = 5.00$. In other words, the synopsis and cast of characters also aided all subjects in reconstructing the story and rejecting false alternatives. The interaction between the synopsis and program was not significant.

For the within-subjects factor of sentence type, the MS_{error} was 225.9. This main effect was significant beyond $p < .001$, $F(5,580) = 13.87$. Hence memory for meaning is in part a function of the type of material tested. The three-way interaction just missed significance ($p < .07$). However, there were significant interactions between sentence type and viewing the program, $F(5,580) = 3.95$ at $p < .002$, and between sentence type and reading the synopsis, at $F(5,580) = 4.38$ at $p < .001$.

The interaction between program and sentence type is illustrated in Figure 1. While memory for meaning is always higher for subjects who have actually seen the program, the improvement beyond guessing levels is more apparent for the clausal items than for either the pronoun-name set or the role-name set. Apparently, subjects who have not seen the program can nevertheless use the internal information in the test booklet itself as well as the information in the synopsis to reject false reference to characters, either in the pronoun-name set or the role-name set. For the clausal items, guessing is close to chance levels. Examination of cell means in Table 1 suggests that guessing on the clausal items is close to random for Group I, while Group II (who have at least read the synopsis) is performing somewhat beyond chance levels.

The interaction between synopsis and sentence type is illustrated in Figure 2. Here too we see that guessing levels are lower for the clausal items than for items involving reference to characters in the story. Perhaps because of these baseline differences, the synopsis has a more dramatic effect on the clausal items. However, there is also a peculiar interaction within the role-name set, such that subjects who have seen the synopsis are better at rejecting false referents for name-target items than for role-target items. For subjects who have not read the synopsis, the reverse is true. These effects are probably related to biases about surface form as well as memory for meaning in the strict sense. Hence these findings should be considered together with the results of the second analysis of variance.

Memory for Surface Form: Analysis over Subjects

Table 2 presents the mean surface memory scores for each of the four groups, for each of the six item types. The analysis of variance on these scores yields a significant main effect of having seen the program, with $MS_{error} = 1880.2$, $F(1,116) = 98.16$, $p < .001$, and a significant main effect of sentence type, with $MS_{error} = 1828.0$, $F(5,580) = 14.54$, $p < .001$. However, in contrast to the

results for memory for meaning, there was no significant main effect of the synopsis on surface memory scores.

There was, however, a significant interaction between program and synopsis, at $F(1,116) = 3.85$, $p < .05$. In general, it appears that the synopsis serves to increase surface memory in those subjects who have seen the program. However, for subjects who are merely guessing which alternative was more likely to occur, the synopsis seems to hinder performance rather than help it. The synopsis and cast of characters may have set off response biases in Group II that differ systematically from the response biases of subjects in Group I, i.e. those who have seen neither the synopsis nor the program.

The three-way interaction was not significant, nor was there a significant interaction between synopsis and sentence type. However, there was a strong interaction between program and sentence type, at $p < .001$, $F(5,580) = 7.89$. Hence, across item types the pattern of surface memory differs significantly from the pattern of response bias in the guessing groups, as illustrated in Figure 3. For example, in the guessing groups there was a clear bias toward elliptical forms vs. expanded clauses. Hence subjects were more likely to obtain correct scores when the target was actually an elliptical clause, and less likely to guess correctly when the target was a full clause. For the pronoun-name set, the biases are in the opposite direction. Subjects were more likely to guess that the referent was a proper name than a pronoun, resulting in higher scores when the target actually was a proper name and lower scores when the target was a pronoun. If we compare this particular bias with performance on memory for meaning on the same pronoun-name items, it appears that the guessing groups have "figured out" who did what to whom, and hence may be anxious to show their knowledge by circling correct names rather than pronouns when given a choice. Finally, among the guessing groups performance on the role-name set averages to chance levels--although the cell means in Table 2 suggest that biases are somewhat different depending on whether subjects have read the synopsis or not.

Figure 3 also illustrates the average levels of performance by item type for those subjects who actually saw the program. However, this particular illustration of surface memory may be somewhat misleading, since it emphasizes deviation from random baselines. Given the finding that guessing biases are far from random among those who have not seen the program, the best illustration of surface memory should contain a correction taking into account the amount of baseline bias that must also be overcome by those who see the program. Figure 4 presents both memory for meaning and memory for surface form in the memory groups, using the standard correction for guessing

$$M_{\text{true}} = \frac{M_{\text{observed}} - G}{1 - G}$$

in which M_{observed} is the average performance of Groups III and IV, and G is the average for Groups I and II.

When the results are plotted in this manner, as in Figure 4, it appears that memory for surface form is better for the more explicit form within each of the three reciprocal item sets. Within the pronoun set, surface memory is better if a name is heard than a pronoun. For the clausal items, there is more surface memory for full clauses than for elliptical clauses. Within the role-name set, surface memory is better for roles than for names. This trend is consistent with the theory of markedness in discourse (e.g. Givon 1975). According to markedness theory, utterances which introduce characters or facts in explicit or marked surface form are generally more novel and informative than utterances that refer anaphorically to referents that are already mentally (if not physically) present in the discourse context. The same effect of markedness does not appear in the results for memory for meaning.

An analyses of variance of the corrected recognition performance scores upon which Figure 4 is based yielded a significant effect for sentence types, $F(4,145)=8.09$, $p<.01$, MS error = .21, as well a significant contrast for unmarked versus marked sentences (A, C, D versus B, D, F), $F(1,145)=4.25$, $p<.05$ for surface memory, but no statistically significant differences for meaning memory, $F>1$. In these analyses, subjects were treated as the random variable. In a parallel analyses with sentences as the random variable the differences in surface memory between sentence types were still significant, $F(5,37)=4.53$, $p = .002$, while the contrast between marked and unmarked sentences did not reach significance, $F(1,37)=2.41$, $.05<p<.10$. No significant differences were found when memory for meaning was analyzed. It should be noted that the analyses over sentences is not very suitable for the design of the present experiment, since the number of sentences per conditions is too small, varying between 2 and 10.

DISCUSSION

The principal findings from this study can be summarized as follows:

(1) Contrary to findings for other types of prose, there is significant recognition memory for surface form for at least these few aspects of dialogue, despite the large amount of material presented in a 20-minute drama.

(2) There are systematic and significant response biases in the two guessing groups, suggesting that "memory" (for meaning or for surface form) should be defined here as the relative gain or overcoming of bias by those who have actually seen the videotaped program.

(3) The type of item tested has a significant effect on both memory for meaning and memory for surface form, in analyses across subjects. In analyses across materials, there is a significant effect of item type on surface memory only. The pattern of results across sentence types is quite different for

surface vs. meaning memory, with a tendency for more marked surface forms to be retained better than less explicit forms.

The first finding is consistent with results by Kintsch and Bates (1977) and Keenan et al. (in press) regarding surface memory for natural language materials, in contexts where subjects do not expect a memory test. Furthermore, in all three studies surface memory is at least in part a product of the type of material tested, and the relationship of that material to the discourse unit as a whole. In the Kintsch and Bates study, surface memory was particularly strong for jokes and other statements that deviated from the "text act" of the lecture itself. In the Keenan et al. study, surface memory was significant only for material high in "interactive value"--a poorly defined dimension, but one that clearly involves the role of an utterance in conversation as opposed to the "memorability" of the individual sentences out of context. In the present study, the effect of item type on either memory for meaning or memory for surface form is quite complex. In general, the amount of surface form retained seems to be affected by the relative explicitness or discourse markedness of a given form--a pattern that does not apply to memory for meaning.

How do we explain the contrast between these findings and reports by Garrod and Trabasso, Sachs and others that surface memory decays within 4-8 intervening sentences or 80 syllables? The difference may be due at least in part to the role of surface form in natural discourse vs. laboratory prose. The distinction between meaning and surface in passages administered out of context, with no real communicative purpose, is indeed a sort of wheat vs. chaff distinction. The surface form has no purpose other than to convey semantic structure, i.e. information about who did what to whom, where, when, etc. Shifts from

active to passive, adverb reorderings, and other paraphrases may have some slight stylistic or rhythmic effect, but in laboratory settings these differences rarely serve the kind of pragmatic function of highlighting or focusing information that they serve in a continuous piece of natural discourse. In natural settings, the surface form may often be the whole point of an utterance-- particularly in jokes and figures of speech. Similarly, in establishing reference in discourse explicit forms are chosen intentionally to draw attention to a referent as either new information, or important and topicalized old information. We suggest that in natural speech, the distinction between meaning and surface form often corresponds to a distinction between semantic vs. pragmatic meaning. The probability that a given surface form will be retained will, at least in part, be a function of the pragmatic role that surface form plays in a given context.

There are some further problems, however, particularly with regard to the distinction between memory-as-retrieval and memory-as-reconstruction. It is possible that subjects in our study have not actually "stored" surface form in the strict sense. Instead, they may use their knowledge of the temporal order of events and settings, and the general flow of the discourse, to decide which surface form a given speaker should have used at a given point in conversation. For example, given the sentence "I wanted to get the Pendleton work done while he was out of the office", a subject will have a higher probability of selecting the correct, pronominal form over the nominal form if he remembers that Robert has already been discussed at length at that point in the conversation. Hence it is more likely that Robert would be referred to by pronoun.

There is a further possibility that subjects are using a mixture of retrieval vs. reconstruction processes, based on the different decay functions for several types of surface information. Recall that in this study (as well as

in the two other experiments on language in natural settings), both visual and auditory information is present in the stimulus materials. This is in notable contrast to the typical laboratory study with written prose. Baggett (1975) has shown that, although there is a semantic integration effect with simple visually-presented stories, the decay function for visually-presented information is much longer than the surface decay for equivalent verbal stories. Indeed, there was some evidence for visual surface memory six weeks after presentation. Similarly, memory for auditorily-presented material is stronger than memory for written versions of the same verbal material (e.g. Murdock, 1974). It is possible that subjects in the present study used bits and pieces of visual and auditory surface memory to reconstruct the surface form of the verbal expression. For example, the explicit vs. anaphoric form in the videotape typically differ in their intonation contours. If the subject can recall a contrastive stress pattern at a given point in conversation (or, alternatively, an emphatic gesture), he may well conclude that the referent was probably encoded in explicit form rather than as a pronoun.

Some of these questions could be answered by further research presenting the auditory portion of the drama alone, and/or presenting the written transcript of the dialogue alone, for comparison with memory for the auditory/visual presentation used in the present study. However, these controls still leave open the possibility that reconstruction on the basis of event chronology is responsible for apparent "retrieval" of surface forms. Undoubtedly reconstruction and retrieval are both used in the processing of natural language materials. To what extent these two processes can be separated empirically remains to be seen.

FOOTNOTE

We are grateful to Herb Clark for discussions concerning the possible role of reconstruction vs. retrieval processes in this experiment.

REFERENCES

- Baggett, P. Memory for explicit and implicit information in picture stories. Journal of Verbal Learning and Verbal Behavior 1975, 14, 538 - 548.
- Bransford, J., Barclay, J. R. and Franks, J. J. Sentence memory: a constructive vs. interpretive approach. Cognitive Psychology 1972, 3, 193 - 209.
- Clark, H. and Clark, E. Psychology and Language. New York: Harcourt, Brace and Jovanovich, 1977.
- Garrod, S. and Trabasso, T. A dual memory information processing interpretation of sentence comprehension. Journal of Verbal Learning and Verbal Behavior 1973, 12, 155 - 168.
- Givon, T. Towards a discourse definition of syntax. Manuscript, Department of Linguistics, University of California at Los Angeles, 1975.
- James, C., Thompson, J. and Baldwin, J. The reconstructive process in sentence memory. Journal of Verbal Learning and Verbal Behavior 1973, 12, 51 - 63.
- Keenan, J., MacWhinney, B. and Mayhew, D. Pragmatics in memory: a study of natural conversation. Journal of Verbal Learning and Verbal Behavior (in press).
- Kintsch, W. and Bates, E. Recognition memory for statements from a classroom lecture. Journal of Experimental Psychology: Human Learning and Memory 1977, 3, 150 - 168.
- Murdock, B. B., Jr., Human Memory: theory and data. Hillsdale, N. J.: Erlbaum, 1974
- Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics 1967, 2, 437 - 444.

LIST OF FIGURES

- Figure 1. Memory for meaning as a function of program by item type.
(Filled circles = program; empty circles = no program)
- Figure 2. Memory for meaning as a function of synopsis by item type.
(Filled circles = synopsis; empty circles = no synopsis)
- Figure 3. Memory for surface form (Groups III and IV) versus guessing bias (Groups I and II) as a function of item type.
(Filled circles = memory; empty circles = guessing)
- Figure 4. Memory for meaning versus memory for surface form as a function of sentence type (corrected for guessing bias)
(Filled circles = meaning; empty circles = surface form)

TABLE 1
Mean Percentage Scores for Memory for Meaning*

<u>Item type</u>	GROUP I (no program, no synopsis)	GROUP II (synopsis only)	GROUP III (program only)	GROUP IV (program and synopsis)
Pronoun Target	83.66%	88.33%	94.66%	97.33%
Name Target	77.66%	81.33%	93.66%	93.66%
Ellipsis Target	68.33%	78.33%	93.66%	93.33%
Full Clause Target	61.13%	71.06%	85.43%	87.90%
Name Target (Role set)	74.93%	87.06%	88.10%	95.10%
Role Target	91.66%	78.33%	93.33%	92.00%

*Chance performance level = 66%

TABLE 2

Mean Percentage Scores for Memory for Surface Form*

<u>Item type</u>	GROUP I (no program, no synopsis)	GROUP II (synopsis only)	GROUP III (program only)	GROUP IV (program and synopsis)
Pronoun Target	-6.33%	-27.00%	-3.33%	-4.00%
Name Target	-0.33%	6.66%	17.00%	29.00%
Ellipsis Target	15.66%	-1.66%	19.00%	30.66%
Full Clause Target	-13.46%	1.60%	24.16%	25.03%
Name Target (Role set)	-8.86%	1.13%	28.90%	39.90%
Role Target	11.66%	-11.66%	66.66%	78.66%

*Chance performance level = 0%







