

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR, 14, 196-214 (1975)

Comprehension and Recall of Text as a Function of Content Variables¹

W. KINTSCH, E. KOZMINSKY, W. J. STREBY, G. MCKOON, AND J. M. KEENAN

University of Colorado

Short English texts, controlled for number of words and number of propositions, but differing in the number of word concepts in the text base (many versus few), were read and recalled immediately. Reading times were longer and recall was less for texts with many different word concepts than for texts with fewer word concepts. Superordinate propositions were recalled better than subordinate propositions and forgotten less when recall was delayed. The probability that a word concept was recalled increased as a function of both the number of repetitions of that concept in the text base and the number of repetitions of the corresponding word in the actual text. These results also obtained when subjects listened to the experimental paragraphs.

Our knowledge about the processes involved in remembering prose lags far behind that about memory for word lists. Inability to represent explicitly the meaning of texts has probably been the restricting factor in this area. There is, of course, more to memory for text than just the memory for its meaning, with important concerns ranging from memory for surface features of a text to the pragmatic aspects of the communication act, but the problem of meaning is a fundamental one. Attempts to bypass this problem, by scoring recall protocols for verbatim recall or ill-defined idea units, have proven to be inadequate and, by their very lack of success, demonstrate the need for a theory-based approach. In recent years there have been several detailed proposals for the representation of meaning which might help psychological research out of this impasse. In the present paper a theory for the representation of meaning proposed by Kintsch (1974) will be used to explore some aspects of memory for text.

The theory assumes that the basic units of meaning are *propositions*. Propositions are n-tuples of *word concepts*, one of which serves as a *predicator*, and the remaining ones as *arguments*, each fulfilling a unique semantic

role. The predicator specifies a relationship among the arguments of a proposition. For instance, in the proposition (LOVE, Experiencer: GREEK, Object: ART) there are two arguments, GREEK and ART, and the predicator LOVE; in English this proposition could be realised with the sentence² *The Greeks loved art*. It is important to note that the arguments of a proposition are concepts rather than words. Word concepts are stored in a person's *semantic memory*, which indicates the combinations of word concepts that form acceptable propositions. Word concepts are defined by the propositions in which they are used. Connected, ordered lists of propositions represent the meaning of a text and are called *text bases*. The ordering of propositions will be discussed further below. In general, however, our interest is not in describing (or justifying) the theoretical system, but in showing that significant experimental questions can be asked within this framework. The extensive examples of text bases which will be presented below should suffice to indicate to the reader the main features of our theoretical notions about the representation of meaning.

Certain properties of text bases may be used as independent variables in psychological

² See Kintsch (1974) for the treatment of tense plurality, and the definite article in this sentence.

¹ This study was supported by grant MH-15872 from the National Institute of Mental Health to the first author.

experiments concerned with the role of meaning in comprehension and retention. One property, number of propositions, has already been studied by Kintsch and Keenan (1973; also in Kintsch, 1974) who showed that reading time increases monotonically with the number of propositions in a text base, even if the length of the corresponding texts is controlled. The present study is a continuation of that work. It includes a replication of the main results and reports on the investigation of the additional variable, the number of word concepts.

Suppose two text bases contain the same number of propositions. Will they be read and recalled equally well? The answer is, of course, that there are numerous other, mostly unexplored factors that affect reading difficulty, and that even if number of propositions in a text base is controlled for, a great deal of variability must still be expected. One factor that may contribute substantially to that variability is the number of word concepts that appear in a text base. For example, a text base consisting of eight propositions may employ only three different word concepts, which are used repeatedly as arguments of different propositions. Alternatively, an eight-proposition text base may use seven different word concepts as arguments, with fewer repetitions. In other words, the former type of text keeps talking about the same few things, while the latter talks about many different ones. Which one is harder to read and remember?

In order to answer this question one would need a fairly elaborate psychological processing model of reading comprehension and memory, which we lack. Informal arguments that one or the other type of materials should be harder are easy to come by. Text bases of a given length with few different arguments are reminiscent of double-function paired-associate lists, in that the same argument is used repeatedly in several different contexts. Thus, interference between the various uses of the argument is predicted for the few argument texts. It should be more difficult to comprehend

such texts, resulting in longer processing times. On the other hand, one can also argue that text bases with many different arguments should be harder. Because the reader must identify each argument and remember its presence in the text base, the more different arguments there are to keep track of, the greater the tax on the limited processing capacity of memory. The finding by Havilland and Clark (1974) that old information in a sentence can be processed more rapidly than new information also indicates that texts with many repetitions of the same arguments should require less time to comprehend than texts with many new arguments.

While we want to obtain empirical evidence concerning this question, there are also a large number of subsidiary problems that will be explored in the present paper. Indeed, Experiments III and IV are mostly devoted to such questions, one to the relationship between reading and listening, and the other to the exploration of long-term retention of prose.

EXPERIMENT I

For Experiment I (and its replication, Experiment II) texts were constructed that controlled the number of propositions in the text base, but varied the number of different word concepts used as arguments of the propositions. Both short and long text bases were employed and in an attempt to keep everything else equal, highly homogeneous sets of paragraphs were used as experimental materials.

Method

Subjects. Thirty-five students from introductory psychology courses who were fulfilling a course requirement participated individually.

Materials. Twelve paragraphs were constructed about widely known topics from classical history. The paragraphs were adapted from "A Child's History of the World," and were, therefore, easy reading material for college student subjects. In each case, a proposition

TABLE 1
MEDIAN NUMBER OF DIFFERENT ARGUMENTS, PROPOSITIONS, AND WORDS FOR THE FOUR DIFFERENT TYPES OF PARAGRAPHS IN EXPERIMENT I (HISTORY) AND EXPERIMENT II (SCIENCE)

Paragraph length	Number of different arguments		
		Few	Many
Short	HISTORY	Arguments	3
		Propositions	8 (7-8)
		Words	8
	SCIENCE	Arguments	21 (21-22)
		Propositions	22 (22-23)
		Words	7
Long	HISTORY	Arguments	8
		Propositions	20 (20-21)
		Words	20 (18-23)
	SCIENCE	Arguments	8 (7-8)
		Propositions	16 (16-17)
		Words	24 (23-26)
	HISTORY	Arguments	66 (64-66)
		Propositions	66 (64-67)
		Words	7 (6-8)
	SCIENCE	Arguments	16 (15-18)
		Propositions	23 (24-25)
		Words	23 (24-25)
	HISTORY	Arguments	70 (67-75)
		Propositions	68 (67-72)
		Words	

Ranges are shown in Parentheses.

list was constructed first, using an actual text passage as a guideline, but without trying to achieve an accurate representation of the text. The main experimental variable was the number of different arguments used in constructing the proposition lists, either *few* or *many*. Factorially combined with this variable was a second factor, the length of the text base, *short* or *long*, as indicated by the number of propositions. Thus, four types of text bases were obtained. Three exemplars of each type were constructed, which were

then translated into English in such a way that the English text expressed everything contained in the text base, and no more. For each type of paragraph, the number of words was kept constant within narrow limits. Table 1 shows the relevant statistics for each of the four types of texts: number of different arguments, number of propositions, and number of words. Every effort was made to produce natural, readable English text (even at the cost of introducing some variability in the number of words and propositions, as is apparent from Table 1). In fact, the experimental paragraphs usually remained quite close to the texts used as guidelines.

In Tables 2-5 examples are shown of the four different types of texts used.³ The semantic cases of the arguments have been omitted, because they are usually obvious. Each proposition is written in a separate row. The rows have been numbered for convenience of reference. Thus, when a proposition is embedded as an argument of another proposition, it is unnecessary to write it out since it can be referred to by means of the correspond-

TABLE 2
THE *Greek art* PARAGRAPH: SHORT, FEW DIFFERENT ARGUMENTS

1	(LOVE, GREEK, ART)
2	(BEAUTIFUL, ART)
3	(CONQUER, ROMAN, GREEK)
4	(COPY, ROMAN, GREEK)
5	(WHEN, 3, 4)
6	(LEARN, ROMAN, 8)
7	(CONSEQUENCE, 3, 6)
8	(CREATE, ROMAN, 2)
Arguments: GREEK, ART, ROMAN (3)	

Text: The Greeks loved beautiful art. When the Romans conquered the Greeks, they copied them, and, thus, learned to create beautiful art (21 words).

³ Copies of all experimental materials are available from the authors.

TABLE 3
THE *Babylonian* PARAGRAPH: SHORT, MANY DIFFERENT ARGUMENTS

1	(BUILD, BABYLONIAN, GARDEN)
2	(BEAUTIFUL, GARDEN)
3	(LOCATION: ON, GARDEN, HILL)
4	(PLANT, BABYLONIAN, FLOWER)
5	(LOVELY, FLOWER)
6	(CONSTRUCT, BABYLONIAN, FOUNTAIN)
7	(DESIGN, BABYLONIAN, PAVILION, 8)
8	(HAS, QUEEN, PLEASURE)

Arguments: BABYLONIAN, HILL, GARDEN, FLOWER, FOUNTAIN, PAVILION, QUEEN, PLEASURE (8).

Text: The Babylonians built a beautiful garden on a hill. They planted lovely flowers, constructed fountains and designed a pavilion for the queen's pleasure (23 words).

ing row number. For instance, the fifth proposition in Table 2 is (WHEN, 3, 4), which is merely shorthand for (WHEN (CONQUER, ROMAN, GREEK), (COPY, ROMAN, GREEK)). Indentation is used to indicate the level of each proposition in the text structure. The theory defines an order relationship among the propositions in a text base: A proposition is said to be subordinated to another if it contains an argument that also appears in the first proposition. With this rule, a hierarchical order of propositions can be specified, given a set of propositions and one or more starting points, the theme (or themes) of the text. (Note that the theory is incomplete in that it has no way of generating the original set of propositions or its theme, for which we rely upon intuition.) Thus, Proposition 2 in Table 2 is subordinated to the first proposition, because it repeats the argument ART. Propositions 3 and 4 are also subordinated to the first proposition because of the repetition of GREEK. Hence these three propositions are assigned to Level 2 of the graph structure (and indented one step). The remaining four propositions are all at the third level of the graph structure (indented two steps), because they all are subordinated to one or more of the Level 2 propositions: Proposition 6 and 8 repeat the argument ROMAN, which was introduced in a second level proposition (num-

ber 3), and Propositions 5 and 7 both embed second level propositions as arguments.

Tables 2 and 3 contain text bases consisting of eight propositions with a total number of 15 arguments in both cases. The difference between them is that only three different arguments are employed in Table 2, but eight in Table 3. Similarly, the text bases shown in Tables 4 and 5 are comparable in length (23 and 24 propositions, respectively) and total number of arguments (39 and 42, respectively), but only seven different arguments appear in Table 4, and 17 in Table 5. There are several other differences between the few-different-argument bases and the many-different-argument bases, which necessarily covary with the few-many difference. First of all, if text bases are constructed from only a few arguments, these arguments must be repeated frequently and there will be more embeddings of propositions as arguments of other propositions. For instance, in Table 2 there are six such embeddings, while only once does a proposition appear as an argument in Table 3. Similarly for the long text we have 15 propositional embeddings in the text base of the *Joseph* paragraph, but only three in the *Assyria* paragraph. Another difference between the two types of paragraphs becomes apparent only if one looks at their graph structures. If one constructs a graph by

TABLE 4

THE *Joseph* PARAGRAPH: LONG, FEW DIFFERENT ARGUMENTS

1	(BECOME, JOSEPH, RULER)
2	(ISA, JOSEPH, SLAVE)
3	(LOC: IN, 2, EGYPT)
4	(RISE, SLAVE, CLASS)
5	(HIGHER, CLASS)
6	(DIFFICULT, 4)
7	(CONJUNCTION, 2, 6)
8	(CONCESSION, 7, 10)
9	(BRIGHT, JOSEPH)
10	(CAUSALITY, 9, 1)
11	(LOC: IN, 1, EGYPT)
12	(GREAT, RULER)
13	(HAVE, JOSEPH, BROTHER)
14	(SEVERAL, BROTHER)
15	(WICKED, BROTHER)
16	(PLAN, BROTHER, 17)
17	(KILL, BROTHER, JOSEPH)
18	(TIME: PAST, 17)
19	(COME, BROTHER, EGYPT)
20	(GET, BROTHER, BREAD)
21	(FINALITY, 19, 20)
22	(FIND, BROTHER, 1)
23	(LOC: IN, 22, EGYPT)

Arguments: JOSEPH, RULER, SLAVE, EGYPT, CLASS, BROTHER, BREAD (7).

Text: Although Joseph was a slave in Egypt and it was difficult to rise from the class of slaves to a higher one, Joseph was so bright that he became a ruler in Egypt. Joseph's wicked brothers, who had once planned to kill him, came to Egypt in order to beg for bread. There they found that Joseph had become a great ruler (66 words).

Note: CONJUNCTION, CONCESSION, and FINALITY are abstract word concepts, realized in the text as *and*, *although*, and *in order to*, respectively.

TABLE 5

THE *Assyria* PARAGRAPH: LONG, MANY DIFFERENT ARGUMENTS

1	(RULE, KING, ASSYRIA)
2	(ISA, ASSYRIA, COUNTRY)
3	(RICH, COUNTRY)
4	(LOC: AT, KING, NINIVEH)
5	(LOVE, KING, LUXURY)
6	(BUILD, KING, PALACE)
7	(WONDERFUL, PALACE)
8	(PLACE, KING, STATUE, ROAD)
9	(LEAD, ROAD, PALACE)
10	(MANY, STATUE)
11	(HUGE, STATUE)
12	(FANTASTIC, STATUE)
13	(DEPICT, STATUE, BULL)
14	(HAVE, BULL, WING)
15	(DEPICT, STATUE, LION)
16	(HAVE, LION, HEAD)
17	(PART OF, HEAD, MAN)
18	(PLANT, MAN', TREE)
19	(MANNER: AS, 8, 18)
20	(RICH, MAN')
21	(TIME, NOW, 18)
22	(LOC: ALONG, 18, DRIVEWAY)
23	(HAVE, MAN', HOME)
24	(LEAD, DRIVEWAY, HOME)

Arguments: KING, COUNTRY, ASSYRIA, NINIVEH, LUXURY, PALACE, STATUE, ROAD, BULL, WING, LION, HEAD, MAN, MAN', TREE, DRIVEWAY, HOME (17).

Text: The rich country of Assyria was ruled by a king who lived at Niniveh. This king loved luxury and he built a wonderful palace. On the roadway that led to the palace he placed many huge and fantastic statues, depicting bulls with wings and lions with the heads of men, just as a rich man nowadays might plant trees along the driveway that leads to his home (67 words).

drawing lines between pairs of propositions that are related via the subordination relationship defined above (that is, they share a common argument), the resulting structures look quite different for few- and many-different-argument text bases. With few-arguments, the network is very tight: Propositions tend to be related to each other all over the list. With many-arguments, on the other hand, each proposition is usually related only

to those propositions immediately following it. That is, there is a steady progression from topic to topic, whereas in the case of the few-argument texts the same concepts keep reappearing (for example, *Joseph* in Table 4).

Procedure. Paragraphs were projected by means of a Kodak Carousel slide projector on a screen about two meters in front of the seated subject. The subject's index finger rested on a response key. Subjects were in-

structed to read the paragraph at their own rate, making sure that they understood it well, and to press the response key when they had finished reading. This removed the slide, and the subject wrote whatever he could remember from the paragraph just read. Instructions emphasized that he recall as much as possible and not worry about whether recall was verbatim. The recall period was as long as the subject required. When the subject indicated that he was finished, the recall sheet was removed and the next slide was presented. The experimental slides were preceded by two warm-up slides, which were read and recalled in the same way as the experimental slides. The order of the slides was random, and was changed for every four subjects.

Scoring. Only recall of complete propositions was scored. Each recall protocol was checked against the propositions of the original text base, and for each proposition it was determined whether it was expressed in the protocol. All protocols were scored independently by two of the experimenters, with an interjudge agreement of 95%. No distinction was made whether recall was verbatim or paraphrase, as long as all elements of the proposition were judged to be present in the recall protocol. All partial recall, inferences, and errors were neglected. In case a subject made an error in one proposition and then repeated that error in a subordinate proposition, the subordinate proposition was scored as correct, to avoid counting errors more than once. For instance, if a subject recalled the *Greek Art* paragraph shown in Table 2 as *The Greeks loved beautiful statues. When they conquered the Romans, the Greeks copied the Romans*, Proposition 1 was scored as incorrect because of the substitution of *statues* for *art*, but Proposition 2 was scored as correct, because the subject remembered correctly that, whatever the object of Proposition 1 was, it was *beautiful*. Similarly, Propositions 3 and 4 would be scored as incorrect in the example given, because of the reversal of *Greeks* and *Romans*. The WHEN-proposition (5) was

scored as correct: Although the embedded propositions 3 and 4 were incorrect, the WHEN-relation did relate the right predicates (CONQUER and COPY). On the other hand, if the recall had been *When the Romans conquered the Greeks, they created beautiful art*, the WHEN-proposition would be scored as incorrect, because in this case *when* relates the wrong propositions; Propositions 2, 3, and 8 would, of course, be scored as correct in this example.

Results

The main results of interest are the reading times and the amount recalled for the four different types of paragraphs. Mean reading times are shown in Table 6. For both short and long paragraphs, reading times were longer for the texts based upon propositions containing many-different-arguments than for texts based upon propositions containing few-different-arguments, even though these texts were equal both in number of words and number of propositions in the underlying text base. An analysis of variance, treating both subjects and paragraphs as random variables, confirmed this conclusion with the *F*-value for the few-many factor being $F(1, 11) = 10.03$, $p = .009$. In addition, the interaction between few-many arguments and paragraph length was also marginally significant, $F(1, 6) = 6.36$, $p = .045$, reflecting the fact that the difference in reading time that was observed between the paragraphs with few and many different

TABLE 6

MEAN READING TIMES IN SECONDS FOR THE HISTORY TEXTS (EXPERIMENT I) AND SCIENCE TEXTS (EXPERIMENT II) AS A FUNCTION OF PARAGRAPH TYPE

Paragraph length		Number of different arguments	
		Few	Many
Short	HISTORY	9.41	10.22
	SCIENCE	9.71	13.92
Long	HISTORY	18.54	23.09
	SCIENCE	29.84	34.53

arguments was much larger for the long paragraphs than for the short ones.

The differences in recall between the few- and many-different-argument paragraphs were much attenuated, but in agreement with the suggestion of the reading times that the many-different-argument paragraphs were the harder ones. Recall was slightly lower for these paragraphs than for the few-different-argument paragraphs, as shown in Table 7, though this difference did not reach statistical significance, $F'(1, 12) = 1.28, p = .28$. Table 7 also expresses the recall scores as percentages of the total possible. It appears that subjects read each paragraph until they were satisfied that they knew enough about it, with a slightly higher criterion for the short paragraphs than

for the long ones (perhaps not making quite enough allowance for the greater difficulty of the many-argument texts), and then tried to recall them. That such feeling-of-knowing can be quite accurate has been demonstrated for list learning experiments (Hart, 1967; Blake, 1973), and there is no reason to suppose that subjects could do less well with texts. Thus, there exists a trade-off relationship between the reading times and the recall scores, making both measures difficult to interpret. A better index of how difficult a particular text is would be a measure that combines both reading time and recall, such as the number of seconds of reading time per proposition recalled. As Table 8 shows, the difference between the texts based upon many- and few-different arguments is quite substantial in terms of this measure. For both short and long paragraphs, a many-different-argument proposition required about .4 second more reading time for recall than a few-different-argument proposition. At the same time, it is interesting to note that the time required to process a proposition was about the same for short and long paragraphs in Experiment I.

There are a number of subsidiary results that are not concerned with the main experimental hypothesis, but that are nevertheless of considerable interest for studies of prose recall. The first one is the plot of number of propositions recalled against reading time, shown in Figure 1, which replicates results reported earlier (Kintsch & Keenan, 1973; also Kintsch, 1974). Figure 1 shows that study time and amount recalled (in terms of number of propositions) are related in a regular manner. The relationship is well approximated by a straight line, though considerably more research would be needed to determine whether, and over what range, linear functions are really adequate. Note that the slope of the functions relating study time and recall is steeper for the long paragraphs than for the short ones (.98 second versus .51 second). This result is also in agreement with the data reported in the original study. Furthermore,

TABLE 7

MEAN NUMBER OF PROPOSITIONS RECALLED (AND PERCENT RECALL) FOR THE HISTORY TEXTS (EXPERIMENT I) AND THE SCIENCE TEXTS (EXPERIMENT II) AS A FUNCTION OF PARAGRAPH TYPE

Paragraph length		Number of different arguments	
		Few	Many
Short	HISTORY	5.44 (65%)	4.79 (59%)
	SCIENCE	6.70 (85%)	4.68 (58%)
Long	HISTORY	11.47 (49%)	11.39 (47%)
	SCIENCE	10.58 (54%)	9.09 (36%)

TABLE 8

MEAN NUMBER OF SECONDS OF READING TIME PER PROPOSITION RECALLED FOR THE HISTORY TEXTS (EXPERIMENT I) AND SCIENCE TEXTS (EXPERIMENT II) AS A FUNCTION OF PARAGRAPH TYPE

Paragraph length		Number of different arguments	
		Few	Many
Short	HISTORY	1.73	2.13
	SCIENCE	1.45	2.97
Long	HISTORY	1.62	2.03
	SCIENCE	2.74	3.80

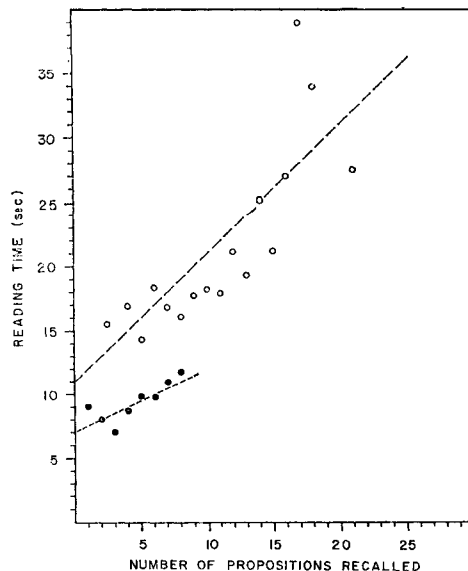


FIG. 1. Average reading time for long (open circles) and short (filled circles) paragraphs in Experiment I as a function of the number of propositions recalled. The broken lines are least square lines.

it is at least in qualitative agreement with corresponding results from list learning experiments, where it is well known that study time increases rapidly with list size (for example, Derks, 1974).

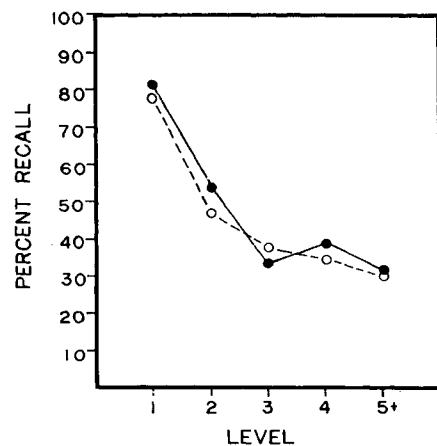


FIG. 2. Recall probability of propositions as a function of their level in the text base for Experiment I (filled circles) and Experiment II (open circles).

A second major result of the Kintsch and Keenan study which is replicated here is the levels effect, shown in Figure 2. Kintsch and Keenan observed that propositions high in the paragraph structure, as defined by the theory, were recalled substantially better than subordinate propositions. In the present study, the level of a proposition, as indicated by indentation in Tables 2-5, was again a powerful determinant of recall, $F(4, 138) = 22.94, p < .001$. Recall dropped from over 80% for Level 1 propositions to about 30% for subordinate propositions. Figure 2 is based only upon the data from long paragraphs. The effect is comparable, however, for the short paragraphs, though there are fewer levels.

In the paragraphs used here there is a tendency for the higher level propositions to be expressed early in the paragraph. Hence there is a possibility that the levels effect may be confounded with a primacy effect. Such a primacy effect would mean that subjects tend to recall the beginning of a paragraph best, no matter what the underlying propositional structure of the text may be. In order to test this hypothesis, it is necessary to evaluate separately serial position in the text and level in the text base. For the purposes of this

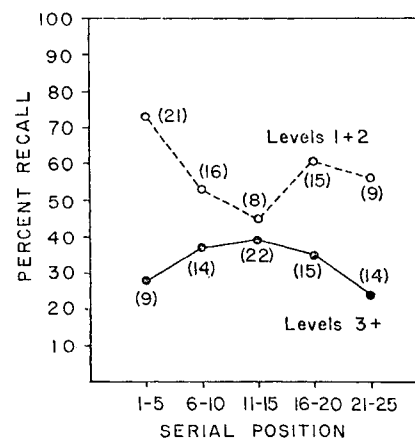


FIG. 3. Recall probability for superordinate (Levels 1 and 2) and subordinate (Levels 3 and higher) propositions as a function of the serial position of their predicates in the surface structure.

analysis, the serial position of a proposition in a paragraph was defined by rank ordering the predicators of all the propositions in the text base according to their position in the actual text. This analysis revealed that superordinate propositions (Levels 1 and 2) and subordinate propositions (Level 3 and below) are, indeed, unequally distributed over the paragraphs, with the former being more frequent in the beginning of the paragraphs (serial positions 1 to 10), and the latter being more frequent in the middle and end portions of the paragraphs (serial positions 11 to 25). The data are shown in Figure 3. The difference is statistically significant, $\chi^2(4) = 12.48$, $p = .014$. However, Figure 3 also shows the recall probabilities for the various combinations of levels in the text base and serial position in the actual text, and it is apparent that superordinate propositions are recalled substantially better at all serial positions. An analysis of variance of the data shown in Figure 3 yielded a significant main effect for levels, $F(1, 133) = 48.76$, $p < .001$, a nonsignificant main effect for serial position, $F(4, 133) = 1.16$, and a significant serial position by levels interaction, $F(4, 133) = 2.81$, $p = .028$. In part, the latter appears to reflect the existence of a primacy effect for Level 1 and 2 propositions, but the complete absence of such an effect for subordinate propositions. Thus, the fact that some piece of information came at the beginning of a paragraph helped it only if it was something important for the text as a whole, that is, a Level 1 or 2 proposition. It should also be noted that there is no evidence of a recency effect in Figure 3, which is, however, not at all surprising since subjects almost always recalled the texts in order. The lack of a recency effect in story recall has also been observed by de Villiers (1974).

In addition to level, other statistics that indicate the importance of a proposition in the paragraph structure were also analyzed, but none was found that influenced recall probability to the same extent as the level of a proposition. For instance, one might suppose

that the total number of connections that a given proposition has with other propositions (two propositions are connected if they share a common argument) might determine its recall likelihood, with propositions with many connections being more likely to be recalled than propositions with few connections. This is to some extent the case, the correlation between number of connections and recall being $r_{RC} = .276$, $p < .001$. But if one partials out levels, this correlation decreases to a value not significantly different from zero. Similarly, if one only counts the number of connections that each proposition has with propositions following it in the text base, the correlation between recall and these descending connections is $r_{RD} = .253$, $p < .001$, which is again reduced to a nonsignificant value by partialling out levels effects.⁴

If one looks at argument recall, rather than propositional recall, the obvious question is what determines recall likelihood: the number of repetitions of a word in the actual text, or

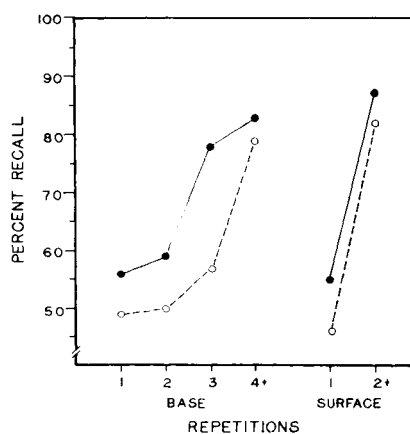


FIG. 4. Probability of recalling an argument as a function of the number of repetitions of the argument in the text base, and as a function of the number of repetitions of the corresponding word in the text for Experiment I (filled circles) and Experiment II (open circles).

⁴ These analyses are reported in more detail in Kintsch (1975); a similar analysis in terms of the number of descendants of a proposition can be found in Kintsch (1974).

the number of repetitions of a word concept in the text base, or perhaps both? For instance, in Table 2 the word *Babylonian* appears only once in the paragraph, but the word concept BABYLONIAN is repeated four times in the text base. A plot of argument recall as a function of number of text base repetitions as well as surface repetitions is shown in Figure 4. Obviously, both types of repetitions have strong effects upon recall likelihood. Of course, the two graphs shown in Figure 4 are not independent, because text base repetitions and surface repetitions are correlated ($r = .53$ for the long paragraphs). Given the design of the present experiment, as well as the ceiling effects apparent in Figure 4, it is not possible to separate the contributions of text base repetitions from the surface repetitions. Theoretically, both types of effects must be expected. If comprehension involves constructing text bases, the more often a concept appears in a text base, the greater should be the likelihood of its recall (though several other factors are probably important, in addition to mere number of repetitions, such as where in a text base a concept is introduced). At the same time, since the task was one of immediate recall, subjects probably remembered surface features of the text in addition to its meaning. Hence, words that were repeated several times in a text should have a recall advantage.

EXPERIMENT II

One obvious characteristic of the texts used in Experiment I was that they were relatively easy (derived from a children's history book), and dealt with familiar topics, though by no means all of our subjects knew about all of them. For Experiment II texts that were both unfamiliar and more difficult were constructed. Articles in *Scientific American* were used for this purpose. Care was taken to select texts only within a fairly narrow field (physical sciences), because pilot studies had shown that it was impossible to obtain a homogenous set of paragraphs dealing with a wide range of

topics. Except for the shift from history texts to science texts, Experiment II closely replicated the first experiment in design, procedure, and analysis.

Method

Subjects. Forty summer school students participated in the experiment. They were recruited from psychology classes and were paid \$1.50 for their participation.

Materials. A total of 16 experimental paragraphs were used in this experiment. Four paragraphs each of the same four basic types as in Experiment I were constructed. The number of different arguments, propositions, and words for each paragraph type is shown in Table 1. An example of each paragraph type is shown in Tables 9-12.

Procedure. Procedure and scoring were exactly the same as in Experiment I, except that the total number of paragraphs was increased from 12 to 16.

Results

The main results of Experiment II replicated those of the earlier study in all critical respects, though there are some interesting differences between the history and science paragraphs. The data are shown in Tables 6-8, along with those for the history paragraphs. Just as in the first experiment, reading times are considerably longer for the many-different-argument paragraphs than for the few-different-argument paragraphs, $F'(1, 13) = 4.71, p = .049$, but the difference is of similar magnitude for both short and long paragraphs, so that the interaction between length and number of different arguments is no longer significant, $F' < 1$. The main difference between the two types of text is that reading times are much longer for the science texts than for the history texts. The average reading time for the history paragraphs is 15.31 seconds, but 21.99 seconds for the science paragraphs. However, subjects made good use of the additional reading time, because as Table 7 shows, recall was just about equal for the difficult science paragraphs (58%)

and the easy history paragraphs (55%). The pattern of recall for the two types of texts is also similar, though the tendency to recall less from the many-different-argument paragraphs than from the few-different-argument paragraphs, which was small and nonsignificant in Experiment I, is more pronounced in the science texts, $F(1, 13) = 5.39, p = .037$. If one looks at the reading time per proposition recalled (Table 8), the main result of Experiment I is again confirmed. Considerably less reading time is needed to recall a proposition if it is part of a text base in which only a few different arguments appear, than if it comes from a text base containing many different arguments. In addition, the reading time per proposition recalled increased with the length of the paragraphs in Experiment II (unlike in Experiment I).

The subsidiary analyses that were performed on the data of Experiment I have also been replicated in the present study. The correlations between reading times and number of propositions recalled are positive for each paragraph, thus replicating the findings of Figure 1. If paragraphs are combined in order to obtain a large enough data base for curve fitting, the slopes of these functions may be estimated as 1.08 seconds for long, few-different-argument paragraphs, and 1.26 seconds for the long, many-different-argument paragraphs, which is somewhat greater than the corresponding value for long paragraphs in Figure 1.

Recall probability as a function of the level of a proposition in the text base is shown in Figure 2. The decrease as a function of level is, again, highly significant statistically, $F(4, 189) = 47.25, p < .001$.

Argument recall as a function of the number of repetitions of an argument in the text base, as well as repetitions of a word in the actual text, are shown in Figure 4. As can be seen, both types of repetitions have strong (though not independent) effects upon the probability of recall. It is interesting to note that argument recall was generally lower for the science paragraphs than for the history paragraphs,

while at the same time propositional recall was essentially the same.

DISCUSSION

The main purpose of the first two experiments reported was to determine whether the number of different arguments in a text base affected reading and recall. The results permit a clear answer to this question. Overall, propositions from the text bases in which only a few different arguments are used required 1.88 seconds of reading time for recall, while propositions from text bases in which many different arguments were used required 2.73 seconds. It is easier for readers to process and retain in memory a proposition that is built up from old, already familiar elements, than to process propositions which introduce new concepts into the text. Propositions that contain new concepts require an additional processing step on the part of the reader: Not only must the proposition itself be inferred from the text and stored in memory as part of the developing text base, but the new concept apparently requires some special processing, in that it must be encoded for memory storage. Old concepts, on the other hand, need not be reencoded; a reference to the already encoded memory trace is sufficient in this case. Theories such as Anderson and Bower's HAM are quite compatible with the conclusions arrived at here (Anderson & Bower, 1973).

Some linguists (for example, McCawley, 1970) prefer a different propositional analysis to the one used here, which makes the distinction between few- and many-different-argument text bases superfluous. Instead, McCawley introduces each argument into a text base by means of a separate proposition, in the manner of standard logic. Thus, the first proposition of Table 9 would have to be rewritten as three different propositions, roughly *X is Mercury*, *Y is field*, and *X has Y*. An existential proposition would also be required for the argument SATELLITE in Proposition 6 of that paragraph, so that the total number of propositions for the *Mercury* paragraph

TABLE 9

THE *Mercury* PARAGRAPH: SHORT, FEW DIFFERENT ARGUMENTS

- 1 (HAVE, MERCURY, FIELD)
- 2 (MAGNETIC, FIELD)
- 3 (NOT, 1)
- 4 (EXPECT, 3)
- 5 (CONTRAST, 4, 6)
- 6 (DETECT, SATELLITE, 7)
- 7 (PRESENT, FIELD)
- 8 (LOC: AROUND, FIELD, MERCURY)

Arguments: MERCURY, FIELD, SATELLITE (3).

Text: Although it had been expected that Mercury would not show a magnetic field, satellites detected such a field around it (20 words).

TABLE 10

THE *Turbulence* PARAGRAPH: SHORT, MANY DIFFERENT ARGUMENTS

- 1 (FORM, TURBULENCE)
- 2 (LOC: AT, 1, EDGE)
- 3 (PART OF, WING, EDGE)
- 4 (GROW, TURBULENCE, STRENGTH)
- 5 (LOC: OVER, 4, SURFACE)
- 6 (PART OF, WING, SURFACE)
- 7 (CONTRIBUTE, TURBULENCE, LIFT, AIRCRAFT)^a
- 8 (SUPERSONIC, AIRCRAFT)

Arguments: TURBULENCE, EDGE, STRENGTH, SURFACE, WING, LIFT, AIRCRAFT (7).

Text: Turbulence forms at the edge of a wing and grows in strength over its surface, contributing to the lift of supersonic aircraft (22 words).

^a The case of LIFT is *Object*, that of AIRCRAFT is *Goal*.

would be 11 rather than eight. To some extent this procedure and the one followed here are merely notational variants, with arguments being introduced implicitly in one case and explicitly in the other. If one reanalyzes the present data counting each argument as a separate proposition (and hence recall of the corresponding word, no matter in what con-

TABLE 11

THE *Asteroid* PARAGRAPH: LONG, FEW DIFFERENT ARGUMENTS

- 1 (ISA, ASTEROID, PLANET)
- 2 (MINIATURE, PLANET)
- 3 (ORBIT, ASTEROID, SUN)
- 4 (IDENTIFY, ASTEROID)
- 5 (NUMBER OF, 4, HUNDRED)
- 6 (CONTRAST, 4, 7)
- 7 (DIFFICULT, 8)
- 8 (KEEP, TRACK, ASTEROID)
- 9 (CONSEQUENCE, 10, 7)
- 10 (ALIKE, ASTEROID)
- 11 (ALL, ASTEROID)
- 12 (IDENTIFY, ASTEROID, POSITION)
- 13 (HAS, POSITION, ORBIT)
- 14 (ONLY, 12)
- 15 (DETERMINE, ORBIT)
- 16 (AFTER, 15, 18)
- 17 (EVEN, 16)
- 18 (LOOSE, ORBIT)
- 19 (CONSEQUENCE, 18, 21)
- 20 (OFTEN, 18)
- 21 (CHANGE, ORBIT)
- 22 (CONSEQUENCE, 21, 23)
- 23 (INFLUENCE, PLANET, ASTEROID)
- 24 (LARGE, PLANET)
- 25 (DEFLECT, PLANET, ASTEROID, ORBIT)^a

Arguments: ASTEROID, PLANET, SUN, HUNDRED, POSITION, ORBIT (6).

Text: Asteroids are miniature planets that orbit around the sun. Hundreds of asteroids have been identified, but it is difficult to keep track of them, since all asteroids are alike. An asteroid is identified only by the position of its orbit. Even after an orbit has been determined, it is often lost because the orbit changes due to the influence of the large planets which deflect the asteroid from its orbit (70 words).

^a The case of ORBIT is *Source*, from which *from orbit* is derived.

text, as recall of that proposition), our conclusions are hardly changed. In either case it is clear that the number of different arguments in a text base is an important determinant of reading time and recall, whether one talks about it in the terms used here or adds that number to the total number of propositions in a text base. Of course it

TABLE 12
THE *Comet* PARAGRAPH: LONG, MANY DIFFERENT
ARGUMENTS

1	(ISA, COMET, FOUNTAIN)
2	(CELESTIAL, FOUNTAIN)
3	(SPOUT, FOUNTAIN, SNOWBALL) ^a
4	(LARGE, SNOWBALL)
5	(FLOAT, SNOWBALL, SPACE) ^b
6	(CONSIST OF, FOUNTAIN, 8)
7	(POSSESS, COMET, 8)
8	(AND, HEAD, TAIL)
9	(SEE, WE, 6)
10	(EXTEND, TAIL, MILE)
11	(NUMBER OF, MILE, MILLION)
12	(CONTRAST, 9, 14)
13	(SEE, WE, SNOWBALL)
14	(NEVER, 13)
15	(HAS, SNOWBALL, DIAMETER)
16	(MEASURE, DIAMETER, MILE)
17	(FEW, MILE)
18	(SHINE, COMET, SUNLIGHT)
19	(REFLECT, SUNLIGHT)
20	(STREW, COMET, DEBRIS, SPACE) ^c
21	(LOC: ALONG, 20, PATH)
22	(HAS, COMET, PATH)
23	(SEE, DEBRIS, EARTH) ^d
24	(APPEAR, DEBRIS, LIGHT)
25	(ZODIACAL, LIGHT)

Arguments: COMET, FOUNTAIN, SNOWBALL, SPACE, WE, HEAD, TAIL, MILE, MILLION, DIAMETER, SUNLIGHT, PATH, DEBRIS, EARTH, LIGHT (15).

Text: A comet is a celestial fountain spouting from a large snowball floating through space. We see the fountain as the head and tail of the comet. The tail extends for millions of miles, but we never see the snowball, which has a diameter of a few miles. A comet shines with reflected sunlight. Along its path it strews debris in space, which seen from the earth appears as the zodiacal light.

^a The case of SNOWBALL is *Source*, from which *from snowball* is derived.

^b The case of SPACE is *Location*, from which *through space* is derived.

^c The case of SPACE is *Goal*, from which *in space* is derived.

^d The case of EARTH is *source*, from which *from earth* is derived.

does make a difference in the estimates of processing times per proposition, as shown in Table 13. If one includes argument recall with propositional recall, recall scores are greatly increased and the estimates of the processing time per proposition recalled are correspondingly lower. In fact, for the history texts these estimates turn out to be approximately equal for the four different types of text used here. This was not the case, however, for the science paragraphs, for which the pattern of results in Table 13 is quite similar to that of Table 8, except that the differences are somewhat reduced in magnitude.

TABLE 13
MEAN NUMBER OF SECONDS READING TIME PER
PROPOSITION RECALLED AS A FUNCTION OF A PARAGRAPH
TYPE, WHERE PROPOSITIONS HAVE BEEN REDEFINED SO
THAT EACH ARGUMENT IS INTRODUCED BY A SEPARATE
PROPOSITION

Paragraph length		Number of different arguments	
		Few	Many
Short	HISTORY	1.11	1.01
	SCIENCE	1.01	1.49
Long	HISTORY	1.06	1.08
	SCIENCE	1.83	1.99

While it is gratifying to note that reading times, recall rates, as well as several other analyses, yielded results in good agreement with the theoretical text structures for both the history and science texts, it is somewhat disturbing that the theory does not account for the large differences in reading times for the two types of paragraphs. Why are the history paragraphs easy and the science paragraphs hard? Conventional readability formulas do not help us very much, in that the two most powerful predictors, word frequency and sentence length, do not discriminate among the two sets of paragraphs. The mean frequency in the Kučera and Francis (1967) word count

of all the arguments in the science paragraphs is slightly larger than that of the arguments in the history paragraphs (123 versus 95), and the average sentence length is practically identical for the two types of paragraphs (15.63 versus 15.56 words). Furthermore, the texts do not differ in abstractness. Using Gillie's (1957) approximation to Flesch's formula, a conventional, if atheoretical, measure of the "abstractness" of texts, an abstraction score of 65 is obtained for the history texts and one of 64 for the science texts. Both texts, therefore, qualify as "fairly concrete." In fact, the only obvious difference between the two types of texts is not a structural one, but a psychological one: The history paragraphs deal with well-known stories, while genuinely new information is provided in the science texts. We have no objective way of substantiating this claim, but it is probably true for most of our subjects. Thus, one may hypothesize that the history paragraphs were easier because readers knew much of what they read already, and their task was primarily one of storing this information in episodic memory. More processing was required for the science paragraphs because a smaller memory base was available relevant to the information processed. The subjects' task with the history texts is somewhat similar to that in a list learning experiment on which he has to learn that certain well-known words were presented as members of a particular list; here he must store the information that certain propositions were presented, but the processing of these propositions is easy because some of them are already part of the subject's general knowledge. Perhaps this is the reason why the reading time per proposition recalled did not differ for the short and long history texts, while a rather sizeable difference (Table 8) was observed for the science texts. In the science paragraphs, propositions had to be newly constructed from the text, a process that becomes more difficult as the length of the text increases and more possible relations among the word concepts need to be considered.

EXPERIMENT III

Subjects read the paragraphs in the first two experiments. In what respects are the obtained results dependent upon this aspect of the procedure? Would there be a difference in the amount recalled, as well as in the type of information recalled, if subjects listened to the experimental paragraphs, rather than read them? Reading permits the subject to return to earlier material at any time and to concentrate his attention upon particularly difficult or interesting passages. This possibility does not exist to the same extent in listening. On the other hand, there are other cues present in listening tasks (for example, intonation contours, pauses) that may help the listener to comprehend the material. The literature on how much people learn from reading versus listening suggests that these opposing factors must be weighted differentially for different subjects and different types of text. There appear to be no consistent differences between reading and listening (Sticht, 1972), but we should nevertheless compare reading and listening with the present experimental materials and tasks, in order to determine how much our results are specific to reading comprehension versus comprehension in general.

In addition, Experiment III permits us to investigate an implication of the results of Experiment I: In that experiment, the many-different-argument texts were read more slowly than the few-different argument texts, but were almost equally well recalled. If reading (or listening) time is kept constant for the two paragraph types, one would expect a greater recall difference in favor of the few-different argument paragraphs.

Method

Subjects. Forty subjects from the same subject pool as in Experiment II participated in this experiment. Subjects were assigned to the reading or listening group alternatively in the order of their arrival.

Materials. The same paragraphs were used

as in Experiment I. All paragraphs were tape recorded by a male speaker on a stereo tape recorder, with natural intonation, and at a comfortable talking speed. The recordings for the short paragraphs were 8 (± 2) seconds long; the recordings for the long paragraphs were 22.0 seconds long (except for one paragraph recorded in 21.6 seconds).

Procedure. For the reading task the procedure was the same as in the previous two experiments, with the modification that reading time was fixed rather than free. The paragraphs were presented for the same amount of time as the tape recordings; that is, short paragraphs were shown for 8 seconds and long paragraphs for 22 seconds. The procedure for the listening task was modified only in so far that tape recordings rather than slides were presented to the subjects.

Results

All of the analyses performed yielded results that were almost identical for the reading and listening tasks. Overall recall was 49% for the reading subjects and 50% for the listening subjects. In Table 14 recall scores are shown separately for the four types of experimental paragraphs. Clearly, the difference between reading and listening is minimal in every case.

TABLE 14
MEAN NUMBER OF PROPOSITIONS RECALLED (AND PERCENT RECALL) AS A FUNCTION OF PARAGRAPH TYPE FOR VISUAL AND AUDITORY PRESENTATION IN EXPERIMENT III

Paragraph length		Number of different arguments	
		Few	Many
Short	READING	5.17 (62%)	4.30 (54%)
	LISTENING	5.58 (67%)	4.53 (57%)
Long	READING	10.56 (45%)	8.93 (37%)
	LISTENING	10.05 (42%)	8.43 (35%)

On the other hand, a comparison of Table 13 with the recall scores after subject-paced reading (Table 7) reveals that the restricted reading and listening data are quite consistent

with the earlier results. First of all, overall recall for the history paragraphs in Experiment I was 55%, which is only slightly larger than the recall level observed here. This is not surprising, since the reading times are not much different: The average free reading time was about 10 seconds for the short paragraphs, and 21 seconds for the long paragraphs, which is quite comparable to the 8 and 22 seconds of reading or listening time in the present experiment. However, when reading times were subject controlled, they were longer for the paragraphs with many-different-arguments than for the paragraphs with few-different-arguments (Table 6), while in the present case both types of short and long paragraphs were presented for the same amount of time. One would therefore expect the more difficult many-different-argument paragraphs to be processed less fully than the few-different-argument paragraphs and hence they should be recalled less well. Table 13 confirms that expectation: On the average, few-different-argument paragraphs were recalled 8% better than many-different-argument paragraphs (compared with a 4% difference when subjects could adjust their reading times freely in Experiment I), $F(1, 8) = 5.05, p = .055$.

More detailed analyses of the data also failed to reveal differences between the recall scores after reading and after listening. If one correlates these scores for each proposition over all the paragraphs, an $r = .86$ is obtained, which shows that the detailed pattern of recall is largely the same after reading and listening. This conclusion is confirmed if one computes the likelihood that propositions are recalled from various levels of the text base: The average deviation between reading and listening functions in plots like Figure 2 is 1%.

It may be concluded that reading and listening were equivalent, and that general features of comprehension and memory are being studied here, rather than reading-specific ones. It is not possible to state on the basis of these results over what range of populations and materials the equivalence between reading and

listening holds. Certainly for poor readers this equivalence breaks down, and it may do so also for longer and more difficult texts, but the present results reinforce the claim made by some reading researchers that the skill of learning by listening is essentially the same and forms the basis for the skill of learning by reading (Huey, 1908; Sticht, 1972).

EXPERIMENT IV

In the preceding studies recall was immediate. In Experiment IV a 24-hour delay was interposed between reading and recalling the texts. It is, of course, to be expected that delayed recall will be much less than immediate recall. The interesting question is which propositions can still be recalled after 24 hours, when memory for the surface features of the texts is presumably quite weak. In the previous experiments we observed that propositions at superordinate levels in the text base tended to be reported more frequently than propositions from subordinate levels. How will this level's effect be affected by a 24-hour delay in recall? Will forgetting be random, that is, will the levels effect not interact with delay, or are superordinate propositions less subject to forgetting than subordinate ones?

Method

Subjects. Twenty-six subjects from the same subject pool as in Experiment II participated in the experiment in individual sessions.

Materials. The history paragraphs from Experiment I were again used in this study.

Procedure. The first part of the experimental session was occupied with a question-answering task that is irrelevant for present purposes. Subjects then read the 12 experimental paragraphs (preceded by two practice texts), and returned for a written recall test 24 hours later. During the study session, the experimental procedure was as in Experiment I, except that no recall tests were administered. Subjects were told to read the paragraphs at their own rate, and that they would be asked

questions about them on the following day. When they returned, they were given a recall booklet with a recall cue for each paragraph printed on a separate page. The recall cue was a statement based on the first, most superordinate proposition of each text base, for example, *The Greeks loved art*, *The Babylonians built a garden*, *Joseph became a ruler*, and *A king ruled Assyria* for the sample paragraphs shown in Tables 2-5, respectively. Subjects were told to recall as much as they could about each paragraph, that recall did not have to be verbatim, but to use complete sentences (in order to facilitate the propositional scoring). In scoring the recall protocols the propositions that served as recall cues were, of course, neglected.

Results

The reading times for the four different types of paragraphs are shown in Table 15. They are slightly shorter than the corresponding results for the history paragraphs in Table 6, but the pattern of results is the same in both cases.

TABLE 15

MEAN READING TIMES IN SECONDS FOR THE HISTORY TEXTS AS A FUNCTION OF PARAGRAPH TYPE IN EXPERIMENT IV

Paragraph length	Number of different arguments	
	Few	Many
Short	8.00	9.16
Long	16.34	20.47

Few-different-argument paragraphs were read faster than many-different-argument paragraphs, $F(1, 5) = 22.56$, $p = .005$, and the interaction of that effect with paragraph length was again present, as in Experiment I, $F(1, 5) = 28.29$, $p = .003$.

Table 16 shows the number of propositions recalled after 24 hours for the different types of paragraphs, as well as percent recall. Compared with immediate recall, these scores are of course much smaller, but they do not differ much for the various types of materials.

The main result of the experiment is shown

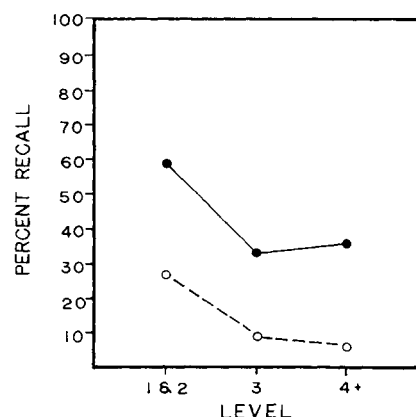


FIG. 5. Recall probability of propositions as a function of their level in the text base for immediate recall (Experiment I, filled circles) and delayed recall (Experiment IV, open circles).

in Figure 5 where recall probability is plotted as a function of the level of each proposition in the text base, both for immediate (same data as in Figure 2) and delayed recall. An analysis of variance reveals a significant levels effect, $F(2, 133) = 8.39$, $p < .001$, a significant delay effect, $F(1, 135) = 417.15$, $p < .001$, and, most importantly, a significant interaction between level and delay, $F(2, 135) = 41.53$, $p < .001$.

TABLE 16

MEAN NUMBER OF PROPOSITIONS RECALLED (AND PERCENT RECALL) FOR THE HISTORY TEXTS AS A FUNCTION OF PARAGRAPH TYPE AFTER A 24 HOUR DELAY IN EXPERIMENT IV

Paragraph length	Number of different arguments	
	Few	Many
Short	1.26 (15%)	1.23 (15%)
Long	4.14 (18%)	3.50 (15%)

One way of expressing this interaction is to calculate, separately for each level, delayed recall as a proportion of immediate recall. Delayed recall was 45% of immediate recall at Levels 1 and 2, but only 27% at Level 3, and 17% at Level 4 and higher. This amounts to a forgetting loss of 55% for the most superordinate propositions, 73% forgetting for the intermediate Level 3 propositions, and 83%

forgetting for the most subordinate propositions. We may therefore tentatively conclude that the level of a proposition in a text base determines not only the likelihood that it will be successfully comprehended by the reader, but also the rate at which it will be forgotten. The conclusion is tentative because the results reported here may depend upon the procedure of cueing delayed recall with a superordinate proposition.

Argument recall after a 24-hour delay showed the same dependency upon number of repetitions as immediate recall. Arguments repeated in the text base once or twice were recalled by 24% of the subjects; arguments repeated three or more times were recalled by 44%. At the same time, words that appeared in the text only once were recalled by 22% of the subjects, but words that were repeated two or more times were recalled by 53%. As pointed out before, these results are inconclusive because the two sets of statistics are not independent, but it is worth noting that even with a 24-hour retention interval such surface features as how often a particular word is repeated in a text may still be very important for recall.

The most striking difference between the immediate and delayed recall results were the massive intrusions that occurred on delayed tests. When recall was immediate, subjects reproduced the text more or less accurately; there were some intrusions, inferences, generalizations, as described by Frederiksen (1972), but they were infrequent and stayed relatively close to the original text. After 24 hours, many of the recall protocols contained more intrusions than reproduction. Differentiation between the experimental paragraph and the subject's knowledge about the subject matter of the paragraph broke down to an astonishing degree. Some subjects seemed to pour forth all they knew about *Babylonians*, *Joseph*, and so on in response to the recall cue. For instance, 27% of the subjects "recalled" that there were "hanging gardens" after reading the *Babylonian* passage (Table 3),

though the term was never mentioned in the original (and none of the subjects used it immediately after reading the paragraph). Sulín and Dooling (1974) have demonstrated a related phenomenon. Their subjects discriminated thematically related distractor items very poorly when given a recognition test after a one-week delay when the topic of the original passage was highly familiar, though no corresponding decline in performance occurred when the topic was unfamiliar.

GENERAL DISCUSSION

The goal of this research was to identify some content variables that significantly affect comprehension and memory for prose. Such data are a prerequisite for psychological models of the comprehension process, which must be based upon a solid empirical foundation. Evidence had been obtained earlier that the number of propositions in the text base that underlies a given text is an important determinant of the rate with which the text can be comprehended and of the amount recalled. A second factor that was identified in the present study concerns the composition of the text base: Text bases that include many different word concepts as arguments of propositions require more processing than text bases with few different word concepts, irrespective of the number of propositions.

Several other phenomena concerned with prose recall were also explored in the present study. Reading time was found to be a function of the number of propositions processed, as indexed by immediate recall. Each proposition processed added a constant amount of time (about one second) to the total reading time.

Predictions concerning which propositions could be recalled after reading a text were quite successful. These predictions were based upon the position of propositions in the structure of the text bases: The more superordinate propositions were recalled better than subordinate ones. This phenomenon was independent of the confounding of the position

of a proposition in the text base with the serial position of the corresponding sentence in the actual text. Superordinate propositions were recalled best at all serial positions in the text. The data also suggested that this levels effect was not only an encoding phenomenon, but that, in fact, forgetting of the superordinate propositions occurred at a reduced rate. The finding that the recall probability of a proposition is closely related to the level of that proposition in the text base structure is similar to one recently reported by Meyer (Note 1). Meyer's definition of levels differs somewhat from the one used here, but the two sets of observations are clearly related.

Recall probability was also analyzed as a function of how often a particular word was repeated in a paragraph, as well as how often the corresponding word concept was repeated in the text base underlying that paragraph. Both types of repetitions appear to be highly effective, even with a 24-hour retention interval, but further research will be needed to separate the two effects.

Finally, memory was compared after reading and listening. When presentation time was the same in both modes, recall was the same after reading and listening for the material and subject populations used here. Thus, the data appear to be relevant to the comprehension process in general, rather than specific to reading comprehension.

In order to reduce the problem of prose recall to manageable proportions, some severe limitations were placed upon the experimental approach in the present studies. The most obvious one is the decision to look at recall only in terms of the correct reproduction of propositions actually stated in the text. Correct inferences made in recall, as well as partial recall, and all types of errors have been neglected. Only the reproductive aspects of recall have been considered here, while its reconstructive aspects, emphasized by Bartlett (1932) and others, have been deliberately excluded from consideration. It should be clear that there is nothing in the theoretical

approach that dictates this choice, but that it was made simply because it was felt that some interesting questions about reproduction could be asked. A theory-based approach that is more concerned with the reconstructive aspects of recall is that of Frederiksen (1972, 1975). Parenthetically, it is of interest that in our studies immediate recall was usually quite accurate, and most errors were omissions. The kind of elaborate intrusions observed by Bartlett occurred only in delayed recall. Naturally, if we used much longer texts reproductive recall would become impossible, and the subject's strategy would shift to one of abstraction-plus-reconstruction.

An interesting problem for further research concerns the interaction between a reader's knowledge base and the text he is reading. There was a large difference in reading times between the easy paragraphs that dealt with well-known topics from classical history and the more demanding paragraphs about scientific topics, about which our subjects possessed little or no previous information. Is this difference only one of encoding difficulty, or does it extend to forgetting rates? Furthermore, is the difference between the history and science paragraphs really one only of pre-knowledge, or are there other, unsuspected factors that need to be taken into account?

The results of the studies reported here provide empirical support for the representation of meaning proposed by Kintsch (1974). The only way to substantiate a theory as complex as that one is by a series of converging experiments. If they can prove their usefulness in other studies, such notions as proposition, argument, and text base may, eventually, lose their speculative character and become accepted psychological concepts.

REFERENCES

- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington, D.C.: Winston, 1973.
- BARTLETT, F. C. *Remembering*. Cambridge: University Press, 1932.
- BLAKE, M. Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon, *Journal of Verbal Learning and Verbal Behavior*, 1973, 13, 311-319.
- DERKS, P. L. The length-difficulty relation in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 335-354.
- FREDERIKSEN, C. H. Effects of task-induced cognitive operations on comprehension and memory processes. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: Winston, 1972.
- FREDERIKSEN, C. H. Acquisition of semantic information from discourse: Effects of repeated exposures. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, in press.
- GILLIE, P. J. A simplified formula for measuring abstraction in writing. *Journal of Applied Psychology*, 1957, 41, 214-217.
- HART, J. T. Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 685-691.
- HAVILLAND, S. E., & CLARK, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 1974, 13, 512-521.
- HUEY, E. B. *The psychology and pedagogy of reading*. New York: Macmillan, 1908.
- KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J.: L. E. Erlbaum Associates, 1974.
- KINTSCH, W. Reading comprehension as a function of text structure. In A. S. Reber & D. Scarborough (Eds.), *Proceedings of the Symposium on Reading*, Brooklyn College, March 1974. Hillsdale, N.J.: L. E. Erlbaum Associates, 1975.
- KINTSCH, W., & KEENAN, J. M. Reading rate as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 1973, 5, 257-274.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.
- MCCAWLEY, J. D. Where do noun phrases come from? In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Reading in English transformational grammar*. Waltham, Mass.: Ginn, 1970.
- STICHT, T. G. Learning by listening. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: Winston, 1972.
- SULIN, R. A., & DOOLING, D. J. Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, 103, 255-262.
- DE VILLIERS, P. A. Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology*, 1974, 103, 263-268.

REFERENCE NOTES

- MEYER, B. J. F. The organization of prose and its effect on recall. Unpublished dissertation, Department of Education, Cornell University, 1974.

(Received November 18, 1974)