JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 19, 96–109 (1980)

Text Structure and Reading Time for Sentences

RANDOLPH K. CIRILO AND DONALD J. FOSS

University of Texas at Austin

Two experiments examining the influence of a story's structure on the comprehension of its sentences are presented. It was expected that sentences at high levels in a story would take longer to encode than those at low levels, either because cues to the sentences' roles exist within the story or because of differential difficulty of integrating the sentences into the prior context. Moreover, the greater density of new information early in stories might result in comprehension being affected by the serial position of a sentence within a story. The reading times for the individual sentences (or clauses) of stories were measured where a particular sentence appeared at one hierarchical (and/or serial) position in one story and at a different hierarchical (and/or serial) position in another story. In both experiments high-level sentences took longer to read than low-level ones and early-occurring sentences longer than late-occurring ones. Recall data supported the structural assignment of the critical sentences. These results were discussed both in terms of the initial hypotheses and in terms of W. Kintsch and T. A. van Dijk's (*Psychological Review*, 1978, **85**, 363–394) theory of text comprehension.

One major focus of recent work on discourse has been to describe the structure of stories and other textual materials (e.g., Grimes, 1975; Kintsch, 1977; Mandler & Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977). Much of this work has involved the development of "grammars" which purport to identify the structural constituents of stories at various levels of organization. Basically, these grammars are organized bodies of rules that represent the multidimensional structure of a story, showing the relationships among the sentences that compose it. It is this underlying organization or "macrostructure" that identifies a sequence of sentences as a story rather than as a random amalgamation.

The existing grammars vary widely both in their complexity and in their range of application, perhaps because they were developed to describe different corpora of textual material (e.g., folk tales, children's stories). However, the grammars also have some similarities; for example, almost all of the representations are hierarchical in nature. At the top levels of the hierarchy are the more abstract components of stories. These are the constituents thought to be typical of stories in general, for example, the setting, the theme. At the lower levels of the hierarchy are the more specific, elaborative details of the story.

The descriptive systems referred to are generally meant to do more than just describe the structure of texts, however. It is usually assumed that when a reader or listener has processed a story adequately, the structure of the story, and not just the structure of its individual sentences, has been extracted from the text. Thus, a description of the structure of a story is also meant to be a (partial) description of the comprehender's representation or knowledge of the story. The description is partial for two reasons: first, the connections among the propositions in the story may need to be expanded upon—by the comprehender making a series of inferences—in order for the representation of the text to be coherent; and second, not all of the struc-

The work reported was supported in part by Grant MH 29891 from the National Institute of Mental Health. Requests for reprints should be sent to Donald J. Foss, Department of Psychology, University of Texas at Austin, Austin, Tx. 78712.

The authors would like to thank Walter Kintsch for a very helpful critique of an earlier version of this paper. Of course, we are responsible for any remaining errors or unclarities.

0022-5371/80/010096-14$02.00/0

ture of the text may be represented in the comprehender's representation of it.

There is evidence that the macrostructure of stories, in particular their hierarchical structure, is a psychologically valid construct. The most compelling evidence is the "levels effect" (Kintsch, 1974; Kintsch & Keenan, 1973; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1978; Meyer, 1975; Thorndyke, 1977). Simply put, the finding is that the hierarchical structure of texts is reflected in their recall. The propositions that stand at the higher levels in the text structure are recalled much better (by about a factor of 2) than those at the lower levels. In general, recall seems to be guided by the comprehender's knowledge about the general structure of stories as well as by the specific information stored in memory about a particular story (e.g., Mandler, 1978).

One of the major aims of the present work was to examine whether the sensitivity to macrostructure exhibited during recall would also be manifest during the course of comprehension itself. It is conceivable that the "height" of a sentence in the hierarchical structure of a story containing it might influence how it is processed when it is first encountered. There are two lines of thought, both fairly vague to begin with, which suggest that a processing difference might exist for high-level vs low-level propositions. The first emphasizes a "top-down" approach and the saliency of cues, indicating that a sentence occupies a high-level position in the macrostructure, while the second emphasizes the difficulty of processing a proposition due to its lack of connections to the earlier text. Both approaches, however, lead to similar predictions with respect to comprehension. Specifically, the first approach argues that the higher a proposition is in the structure, the more salient it is for the reader. Both Bower (1976) and Mandler (1978) have stated that their story frameworks would allow readers to distinguish major from minor events. Further, Bower has suggested that the high-level propositions probably receive more attention since they fill the basic, essential slots of the story frame. Extending this suggestion, one might propose that the reader is guided by his or her representation of the story framework to process more thoroughly—that is, take more time with—those sentences that play a high-level role in the framework. Naturally this proposal requires a processing system that can detect rapidly which propositions are high level. Such a view suggests, though it does not require, that there are cues in the text which signal to the reader the level that a sentence occupies in the story structure.

Alternatively, it is possible that high-level propositions are more difficult to integrate with the previous context as the overall macrostructure is being built during comprehension. High-level propositions typically introduce new material rather than expand upon material already presented. Thus, there are few bridges (Clark, 1977) between the concepts involved in the high-level proposition and the concepts already introduced in the text. The high-level propositions can be seen as standing at the writer's major choice points, where he or she is most free to take the text in new directions. In this sense, the content of the high-level propositions may be less predictable than the content of the lower-level ones which tend to elaborate upon already established ideas. As a result, one would again expect that high-level propositions require more processing than low-level propositions, although in this case the processing difference would not necessarily lead to the levels effect in memory.

To date there have been few attempts to account for the levels effect in memory or to state how sentences in texts are differentially processed as a function of their position in the text structure. One notable exception is the recent work of Kintsch and van Dijk (1978). They proposed a processing model that does attempt to account for the levels effect in memory. Their model concerns itself both with the global macrostructure of the text—that which characterizes the discourse as a whole—and with its more local microstructure, "the structure of the individual propositions and their relations" (p. 305). According to their model, the input sentences are first represented as propositions (relational concepts plus their arguments). In order to discover and form in memory a coherent representation of the text, the propositions are then checked for referential coherence, that is, overlap of referents among the propositions. Specifically, according to the (microprocessing) model, the initial *n* propositions of a story are entered into the system one at a time during a processing "cycle," and a subset of them is selected and stored in a short-term buffer memory of limited capacity. During the next processing cycle, when the next *n* propositions are entered, the propositions in the buffer are available for connecting the new material with the old. The text is coherent if an overlap can be found between arguments of the new propositions and arguments of those propositions stored in the buffer. The model permits a proposition to continue to reside in the buffer during numerous processing cycles. The longer a proposition remains in the buffer and the more times it is processed, the higher is the probability that it will be stored in and reproduced from long-term memory. Kintsch and van Dijk speculate that high-level propositions typically remain in the buffer during more processing cycles than do low-level propositions. Therefore, high-level propositions have a higher probability of being stored in long-term memory and of being recalled. Note that this aspect of the model assumes that the advantage in memory gained by a high-level proposition takes some time (processing cycles) to accrue. If this were the only mechanism available to distinguish high-level from low-level propositions in Kintsch and van Dijk's model, then the model would predict that there would be *no* immediate difference in comprehension time required for high- vs low-level propositions.

There are, however, two other features of Kintsch and van Dijk's model that could differentially affect high- and low-level propositions. The first still has to do primarily with the level of microstructure. As noted, Kintsch and van Dijk state that the reader or listener looks for overlap between the arguments in the new propositions, those read in during a cycle, and the arguments of the propositions stored in the buffer. If no such overlap is found, then a resource-consuming search of long-term memory is made to see whether there are arguments from earlier propositions, no longer in the buffer, that overlap with those from the new input. This aspect of their model is similar to the point made earlier when we noted that sentences in a text may be more difficult to process when bridges are missing between them and the earlier material in the text. Such bridges are more likely to be absent for high-level sentences. Hence, degree of overlap among arguments could be a relevant variable in the processing of high- versus low-level propositions, one that might lead to differences in the time needed to comprehend. Another relevant feature of the Kintsch and van Dijk model is related to the processes associated with the macrostructure. There are, in the model, a set of macrooperators which "transform the propositions of a text base into a set of macropropositions that represent the gist of the text" (p. 372). These operators are guided by the reader's goals in reading, and they operate probabilistically on the propositions. Similar to the earlier part of the theory, macrooperations apply in cycles; a macroproposition is more likely to be recalled the more cycles it is processed. In other words, high- and low-level propositions must be differentiated by the macroprocesses if the gist of a text is to be formed. Therefore, it is possible that such processing could have immediate effects (of the sort we have hypothesized) on comprehension. However, since the macrooperators are controlled by the reader's goals or schema, and since these latter concepts are as yet relatively vague, this part of the model is not as explicit as is the earlier, microstructural segment and it does not specifically predict such effects. The work to be reported below bears upon the Kintsch and van Dijk model, although our work was under way before their paper appeared and thus was not explicitly designed to be a test of their model.

To reiterate, we expect that a sentence will require more processing if it is "high" in the hierarchical structure of a story than if it is "low" in the structure. This will happen, we think, either because high-level sentences are cued as such, or because high-level sentences are more difficult to integrate with the preceding context. Consequently, an on-line measure of comprehension should reveal differences among the components of a story depending upon their level in the overall story structure. The present studies will examine this hypothesis by measuring the reading times for the individual sentences or clauses of simple stories and folk tales. Differences in processing requirements for a sentence are expected to show up as differences in reading time. While this is probably not the most satisfying measure of processing load, it does permit relatively natural processing of the material and is thus a reasonable place to begin an examination of the course of story comprehension.

In order to isolate the structural effects of current interest from a host of potentially confounding variables that might affect reading times for sentences (e.g., intrasentential differences like different syntax or lexical items), single sentences were taken from structured stories and placed in new stories at a different level in the hierarchical structure. Thus, we can compare reading times for identical sentences which play different roles in their respective stories. If a sentence takes longer to read in one story than in another, it is likely that the difference is due to differential processing of the sentences required because of their different levels in the story structure. Specifically, if the above considerations are on the right track, we predict that a sentence at a high level in a story structure will take longer to read than will the same sentence at a low level.

Determining the structure of texts is, at present, a far more subjective enterprise than is desirable. We used an empirical measure of height in a text to validate the structural analyses posited for the stories that were used. We noted earlier that high-level propositions tend to appear in recall protocols more often than do low-level propositions (the levels effect). Accordingly, recall data were collected in the present experiment. If the hypothesized story structures used in these experiments are correct, then the critical sentences should appear more often in the recall protocols of stories in which they play a high-level role. The same sentences should be recalled less often when they play a low-level role in a story.

The present experiments were also designed to validate an aspect of the ongoing processing of stories suggested by pilot work using the reading-time technique. It appeared to be the case that the serial position of a sentence within a story exerted an influence on its reading time. A sentence appearing early in a story seemed to take longer to read than the same sentence appearing later on. Such an effect is not surprising when one considers the information content of a story as a function of serial position. If most new information occurs relatively early, then later information will be easier to integrate and hence faster to process. Of course, a general trend toward faster reading times as one gets into a story (if there is indeed such an effect) would be subject to much fluctuation since local variables as well as more structural ones no doubt influence the reading process.

Experiment I

The first experiment attempted to investigate the hypothesized effects of hierarchical level and serial position on reading time. This was done by using two sets of stories. In one, the hierarchical level of a sentence was varied across stories while serial position was held constant. In the other set, serial position was varied while hierarchical level remained constant. In the former set, high-level sentences should take longer to read than low-level sentences, while in the latter set, early-occurring sentences should take longer to process than late-occurring sentences.

Method

Subjects. The subjects were 21 undergraduates from the University of Texas at Austin. They participated in the experiment as part of a course requirement.

Materials. A base set of 10 stories was obtained for use in the experiment. Several were taken from previous research in prose comprehension (e.g., Thorndyke, 1977). The remainder were found in a collection of Armenian folk tales. The only criteria for selection were that each story be reasonably comprehensible (i.e., no odd narrative structures as in "The War of the Ghosts" used by Bartlett, 1932) and relatively short so that a number of them could be presented during the allotted time. Some of the stories were edited down because of length considerations.

From this base set, two experimental sets of stories were constructed. For the first set, five of the base stories were taken and each was divided into units consisting of simple sentences or clauses. Thorndyke's (1977) story grammar was then used to derive hierarchical structures for the stories. Using these structures, one of two methods was used in order to generate five new stories. In the first, one of the units (always a complete sentence) which occurred at a high level in its particular structure was extracted. It was then inserted into a new story in which it would fit at a low level, without distorting the story line. In the second method, one of the units which occurred at a low level in its structure was extracted. It was then inserted into a new story in which it fit at a high level. In both methods, new stories were either written from scratch or found in the collection of folk tales. In manipulating the extracted units, their position across stories was held approximately constant. If the critical unit occurred early (late) in the base story, it occurred early (late) in the new story. "Early" and "late" refer here to the first and second halves of the stories, respectively. Thus it was possible to test for an effect of hierarchical level on the processing of the critical units apart from any effect of serial position. The critical units, then, were paired across two of four conditions: high level–early occurrence with low level–early occurrence, and high level–late occurrence with low level–late occurrence. An example of the paired stories is shown in Tables 1 and 2.

It should be noted that due to problems inherent in dealing with stories, a constant distance between high and low levels could not be maintained when new stories were generated. However, "high-level" units were always in the first or second levels of their particular structures, while "low-level" units were separated by at least one level in their story structure from that occupied by their corresponding "high-level" unit.

The second set of stories was generated from the remaining base stories by methods similar to those used for the first set. For these stories, however, the critical units maintained their structural height across story pairs while serial position was varied. Thus high level–early occurrence was paired with high level–late occurrence, and low level–early occurrence was paired with low level–late occurrence.

Because of limitations engendered by the mode of presentation, the sequence of units

TABLE 1
EXAMPLE STORY IN THE HIGH–EARLY CONDITION

The Witch's Curse

There was once a far off land
which was ruled by a wise king.
One day a woman was brought before him.
She was accused of witchcraft.
The evidence was quite clear,
and the king ordered her death.
Hearing this, the witch cursed the king.
The king soon felt the curse's effect.
He could no longer talk at all.
No sound came from his throat.
The king quickly decided that
this curse must be lifted,
if he was to rule effectively.
He put the task to his magician
who called upon the powers of light
to come to his aid.
Soon, a whisper was carried to him
by the gusts of a gentle breeze.
It told the magician that
there was a great evil on the land
which they were powerless against.
Thus, the magician had no choice.
He had to turn to the powers of dark.
He muttered many incantations,
and a voice came out of the shadows.
It agreed to perform the task for him
in return for a soul of its choosing.
The magician agreed to this,
fearing the anger of the king.
That morning, the king was found dead.
Every night, his voice is heard,
echoing through the corridors.

TABLE 2
EXAMPLE STORY IN THE LOW–EARLY CONDITION

The King's Ring

Long ago, a king lost his ring
while on a drive through his capitol
A simple foot soldier found it.
He heard about the large reward.
He was shocked by his good fortune.
He could no longer talk at all.
He soon recovered, however.
He decided to take the ring to the king
so he could get the large reward.
Arriving at the gate of the palace,
the guard refused to let the soldier in
unless he gave him half the reward.
This angered the soldier greatly,
but he had no choice but to agree.
The soldier, however, asked the guard
to give him a note stating that
they were to split the reward.
Upon the return of the valuable ring,
the king offered the soldier much money
but the soldier stated solemnly that
a soldier's reward is 200 lashes.
Thinking the man to be a fool,
the king ordered the lashing be done.
But the guard was due half the reward.
So as the soldier undressed,
he presented the note to the king.
After the king had read it,
he had the guard given 100 lashes.
As the last 10 lashes were given,
the soldier drew near to the king
and, noting the guard's apparent greed,
offered to give him all of the reward.
Immediately, the king ordered this too.
After this large and generous reward,
the guard could hardly crawl home.
As for the simple foot soldier,
the king gave him a great deal of gold
which he gratefully accepted in full.

that comprised the stories was edited such that each unit was no longer than 40 characters. As stated before, units were not always complete sentences. However, the critical units *were* always complete sentences. Moreover, they were embedded in a context of complete sentences so that they would be treated in a normal fashion. The stories in Tables 1 and 2 are in the final edited form. The stories ranged in length from 25 to 39 units, the average being 33.1.

Procedure. Subjects participated in the experiment singly in a darkened booth. The stories were presented one unit (sentence or clause) at a time on a Tektronix Display Scope controlled by a DEC PDP-8I computer. The subjects sat approximately 1 m from the screen. The display was green against a dark background; the letters subtended about 12 × 15 min of visual angle. Subjects controlled the rate of presentation by pressing a button which eliminated the unit currently on the screen and immediately replaced it with a new unit. The computer measured the time between button presses to the nearest millisecond, providing a set of reading times for the units of a story.

Each unit was left justified on the screen. New units appeared directly below the positions from which old units disappeared until the bottom of the screen was reached, at which point presentation shifted to the top of the screen again. Subjects were informed as to the manner of presentation before they began reading.

The screen was blank between stories. In order that the start of a new story would not be too abrupt, stories were preceded by the sentence "Push button when ready for paragraph." Subjects were instructed that this phrase signaled the start of a new story and that they should only proceed when fully prepared to do so. Subjects were also instructed to read the stories at whatever pace they found most comfortable. They were told, however, to read each story carefully enough to comprehend it well, keeping in mind that a comprehension test would be administered after all the stories had been read.

Subjects read the first set of stories (those in which structural height was varied while serial position was held constant). The order of presentation was controlled by the experimenter in order to minimize the probability that subjects would notice the repetition of the critical sentences. Paired stories were always separated by four of the remaining experimental stories. Thus the order was not completely random, but every story did appear in every position across subjects. Moreover, for every subject (except one) who saw one member of a story pair first, there was another who saw the other member first. No subject questioned after the experiment mentioned noticing the repetition.

After the first set of stories, there was a short break of approximately 5 minutes immediately followed by the second set of stories (those in which serial position was varied while structural height was held constant). The procedure was exactly the same for these stories.

Following the second set, subjects were asked for a free recall of the last four stories they had read. They were instructed to make their recalls as close to verbatim as possible, but not to leave anything out because they were unsure of wording. Subjects were cued as to which stories to recall by being given titles to the last four stories they had read. The subjects had not seen the titles when they first read the stories, but the titles were assumed to be uniquely identifiable with the story of interest. Cues were deemed necessary because of possible confusions due to the large number of stories the subjects had read. Having been given the cues, the subjects were told to work on their recalls one at a time in the order in which they were given (least to most recently read). Subjects were also told not to go back to a story once they had finished working on it. No time limits were placed on the recalls. The finished protocols were scored for the presence or absence of the *gist* of the critical sentence present in the story they read. Only the second set of stories was tested in this fashion since the structures for the first set were validated in pilot work.

Results and Discussion

The mean reading time for each condition (high–early, high–late, low–early, low–late) was computed for each subject. In compiling these data, the times for certain critical sentences were discarded due to a problem with the computer program used to control the presentation. Specifically, during the course of the experiment units were sometimes skipped if the button was held down too long. It was feared that such skips would disrupt the processing of the material. For this reason, if such a skip occurred at *any* time prior to the presentation of one of the critical sentences, its reading time was discarded. In order to control for the possible effects of particular sentences, the time for the corresponding sentence in its paired story was also discarded in such cases. As a result, 5% of the total number of reaction times collected in this experiment were not included in calculating the following statistics.

The results of each condition for the height manipulation are shown in Table 3. The pattern of reaction times supports the

TABLE 3
MEAN READING TIMES (msec) FOR CRITICAL SENTENCES: HEIGHT MANIPULATION

	Serial position	
Hierarchical level	Early	Late
High	1854	2055
Low	1770	1574

TABLE 4
MEAN READING TIME (msec) FOR CRITICAL SENTENCES: SERIAL POSITION MANIPULATION

	Hierarchical level	
Serial position	High	Low
Early	1833	1643
Late	1578	1632

hypothesis that the structure of a story influences the processing of its components. High-level units took longer to read than low-level units. An analysis of variance by subjects showed a significant main effect for height in a story structure, $F(1,20) = 10.16$, $p < .005$. A significant interaction between height and position was also found, $F(1,20) = 4.61$, $p < .05$. It is difficult to say whether or not this interaction reveals something about processing or is only a product of the materials. The problem is that, although stories are paired across hierarchical position, they are not paired across serial position. Thus, the high/low difference obtained for the early position is a product of different sentences than that obtained for the late position. Thus, some differences among the critical sentences themselves could have led to the interaction. In fact, one of the story pairs in the late condition resulted in an abnormally large high/low difference in reading time relative to the other pairings. This could have been the locus of the interaction. Even though this materials confounding makes interpreting the results more difficult, the data are still supportive of our initial hypothesis.

Table 4 shows the results of the early/late manipulation. Here a significant main effect of serial position was found, $F(1,20) = 4.51$, $p < .05$, as well as a significant interaction between serial position and height, $F(1,20) = 5.13$, $p < .05$. The early parts of a story took longer to read than the later parts, although this effect appears to be restricted to the high-level units. Again, one must note that the obtained interaction is problematic for the same reasons cited earlier in discussing the data from the height manipulation. Differences in the critical sentences could be influencing the size of the early/late effect in reading time.

The scoring of the recall protocols supported the structuring given to the stories. As one would expect given the previous results from studies dealing with the recall of texts, the high-level critical sentences appeared in the protocols much more often than did the low-level units: 83 and 15%, respectively.

The results from the first experiment are promising. They show that the reading-time technique can yield data that are reliably affected by such manipulations as hierarchical and serial position within stories. However, the experiment can be faulted because it was not truly factorial in design. Sentences were paired across only two of the four conditions created by varying both hierarchical and serial position. As a result, it was impossible to test adequately for the effects of the two variables simultaneously. The apparent interactions could have been due to materials effects. It is important to determine whether or not the interactions were due to the two variables of interest, hierarchical and serial position. Different interpretations of the hierarchical effect will be required if the effect interacts with serial position than if it does not.

EXPERIMENT II

A second experiment was designed in which the critical sentences were repeated across four stories, varying hierarchical and serial position orthogonally. Even though the large number of repetitions might be expected to have some influence on processing, it was expected that the effects of hierarchical and serial position would be maintained.

Method

Subjects. Subjects were 24 undergraduates from the University of Texas at Austin who participated as part of a course requirement.

Materials and procedure. Three pairs of stories were retained from Experiment I. For each of these pairs, two new stories were written (or old stories rewritten) which utilized their critical sentences in positions which filled the remaining cells of the 2 × 2 matrix. The result was three sets of four stories. The critical sentence in each story fulfilled one of the four experimental conditions: high–early, high–late, low–early, low–late. These stories comprised the test set. They ranged in length from 27 to 38 units, the average being 32.3. Since pilot work and Experiment I indicated that the effects of interest appeared in even the first stories read, a practice set of only four stories was used in the interests of time. The practice stories were culled from those dropped from the earlier experiment. The number of stories recalled was also reduced to three. In all other respects the procedure was the same as that of Experiment I.

Results and Discussion

Table 5 shows the reading-time data from Experiment II. Four percent of the RTs were excluded from consideration due to problems of "skipping"; see Experiment I. Two additional times of 5.83 and 10.13 seconds were deleted (along with their associated times) as being abnormally long in comparison to the subjects' other reading times. As in the previous experiment, an analysis of variance showed significant effects of both hierarchical position, $F(1,23) = 12.15$, $p < .005$, and sequential position, $F(1,23) = 6.12$, $p < .025$. There was no interaction between these two variables.

TABLE 5
MEAN READING TIMES (msec) FOR CRITICAL SENTENCES: EXPERIMENT II

	Serial position	
Hierarchical level	Early	Late
High	1966	1785
Low	1837	1602

The main effects observed here are quite similar to those found in Experiment I. Thus, they provide additional evidence for the effects of story structure and serial position on the processing of stories. Also, as in the earlier experiment, the recall data supported the structures used for the stories. Critical sentences were recalled 66% of the time when they were high-level units as compared to only 26% of the time when they were low-level units.

GENERAL DISCUSSION

The results from the two experiments support the hypothesis that readers are sensitive to the structure of a story as they read it. Subjects took longer to read a sentence when it played a high-level role in a text than when the same sentence played a low-level role, and they took longer to read a sentence when it occurred early in a text as compared to when it occurred late. Thus, the "levels effect" appears to hold for sentence processing as well as for recall.

Earlier we noted that Kintsch and van Dijk accounted for the levels effect in memory by reference to the amount of time (number of processing cycles) that a proposition remained in the buffer memory. While this may be a correct interpretation, the present results call it into question. Readers do not spend equal amounts of time processing high- and low-level propositions when they are first encountered. Thus, the memory results may be affected by this initial processing difference rather than (or in addition to) the difference in time that items spend in a buffer memory. In the present experiments the subjects were asked to recall only the last three or four stories they read, so a comparison of recall scores with reading time as a covariate is not practical. Inspection of the data did suggest, albeit slightly, that reading times were longer for those critical sentences that were recalled than for those that were not recalled. This is a matter for further systematic investigation. At a minimum, the present results suggest strongly that investigators should pay attention to the immediate processing of sentences if an adequate theory of text memory (let alone text comprehension) is to be constructed.

Let us consider first the influence of hierarchical position on reading times. To reiterate, we began with two broad hypotheses which led us to expect that readers might, in fact, be sensitive to this aspect of text structure during comprehension. We suggested first that subjects might be able to determine which propositions of a story are important (and therefore should be processed more thoroughly) as they are encountered because text processing is guided by story frameworks or schemata, and because the texts may contain cues concerning the role that the various sentences play. Alternatively, we suggested that high-level sentences might require more processing time because they are more difficult to integrate into the existing structure that the reader is building.

According to the first broad hypothesis, the one suggesting that processing is directed by the schema and by cues to the macrostructure present in the text, readers know that certain units are more important than others. They use these expectations to guide their processing of new texts, expectations which are no doubt themselves guided to a large extent by the reader's past experience with events and their descriptions. In consequence, they "intend" to devote more processing time to the important units in order either to integrate them well with other information in memory or to help make sure that these units are stored in long-term memory. Of course, if the present data are to be accounted for by an appeal to this hypothesis, it is necessary that the processes associated with the macrostructure act upon the immediate processing of the sentence.

However, such a schema-based processing model presents a number of difficulties. One theoretical problem concerns how best to describe reader expectations. Another has to do with how the expectations enable the reader to develop procedures for analyzing new events or their descriptions. And a third concerns how the reader determines which schema or expectations are relevant during text processing.

Although each of these problems will require study, they cannot all be considered here. Instead, let us note an aspect of the first problem which arises out of our use of a story grammar as an analytical tool in the present experiments. Macrostructures are often said to be generated by story grammars. It is sometimes not clear exactly what claims are being made concerning the ontological status of such grammars. Are they appropriate descriptions of the reader's expectations? Are they meant to be incorporated somehow into psychological models of text comprehension? Or are they simply claims about the structure of the text itself, a distant reflection of the mechanisms that produced the structure? Given the fact that macrostructural effects are manifest in the reading times, one might be tempted to conclude that the rules of a story grammar are psychologically valid in the sense that they are directly utilized in comprehension. Such a "direct incorporation" model would bear a close resemblance to models of sentence comprehension that tried to incorporate directly (via analysis-by-synthesis models) the rules of sentence grammars. However, since such models have been found wanting both theoretically and empirically (see Fodor, Bever, & Garrett, 1974, for a review), it seems highly unlikely that they would be adequate at the level of stories. In addition, the existing story grammars are not, technically speaking, grammars at all. They do not specify adequately either the terminal or the nonterminal symbols of their vocabularies, let alone the rules; they do not meet even the most minimal standards of descriptive adequacy.

Whatever the case, text structures are quite flexible. As a result, top-down schema-based models would probably have to rely upon particular cues in the story to aid readers in assigning appropriate structures. Both Kintsch (1977) and Winograd (1977) hold that certain cues in the text, together with the current context, can trigger a hypothesis that an instance of some schema or subschema is being conveyed, completed, or replaced. It is also possible that there are some cues concerning the position in hierarchical structure that a proposition is to occupy. If so, then reading-time differences of the sort observed in Experiments I and II would be predictable. Of course, such an account relies on there actually being regular relations between the proposition being processed and its context which can serve as cues to the proposition's role in the structure. At the present time the nature of these cues is simply a matter for speculation. Among other possibilities, such cues could take the form of shifts in subject or in verb tense, the type of connection between the current sentence and those preceding it (e.g., a temporal sequence versus causal implications), the presence of a referring expression that points to an already important referent, and so on. As Olson, Duffy, and Mack (in press) have suggested, however, some propositions in a text may seem important when first encountered even though they are not central to the final representation. Thus, the cues we are hypothesizing should only be considered relative and not absolute. Considerable work will need to be done to remove this point of view from being mere speculation.

The second broad hypothesis concerning the levels effect in reading times says that high-level propositions take longer to read because they are more difficult to integrate into the structure that the reader is building. This view, which is closer to a bottom-up perspective, assigns a considerably reduced role to the construct of "expectation." Within this framework, readers and listeners do not have to be credited with sets of expectations about what is to come; the role of story grammars is thus somewhat reduced.

The problems of integrating propositions with earlier context may be clarified somewhat by considering the following sequence from Table 1, representing a downward transition in hierarchical level: "He could no longer talk at all. No sound came from his throat." Having set up the condition that the king could no longer talk, one can easily accept the fact that no sound came from his throat. The second sentence is clearly a consequence of the first and, as such, may be readily integrated into the macrostructure that is being built by the subject. Therefore, it should require little reading time. The same is likely to be true for most of the propositions that Thorndyke (1977) views as directly appended (at lower levels) to other propositions in a story: topical elaborations, further specifications, or causal implications.

Now consider the following sequence from the same story, representing an equivalence in hierarchical level: "There was once a far off land which was ruled by a wise king. One day a woman was brought before him." It is obvious that the second sentence here does not relate to the previous context as well as did the second sentence in the previous example. In the present case the second sentence does not elaborate upon the first, nor is its truth apparent given the truth of the first sentence. As a result, it should not be much easier to integrate into the whole than the first sentence itself (save for the anaphoric connection between "king" and "him") and hence should take at least as long to read, all else being equal.

Note that in both of the example pairs there is one argument in common between the two sentences: "He–his" in the first example, and "king–his" in the second. In these cases, then, the pairs are equally related to each other according to the metric provided by Kintsch and van Dijk. Recall that these theorists proposed that new sentences would be readily integrated into previous text if there was an overlap in arguments between the input proposition and the information in the memory buffer. (For purposes of this one illustration, the buffer is assumed to contain the previous proposition.) Thus, the second sentence of each of the above pairs would not require the differential processing our own considerations suggest. In order to examine this aspect of Kintsch and van Dijk's model and its possible relation to the present results, the materials from Experiment II were inspected to see whether there was a difference in overlap among arguments between context and the high- versus low-level sentences. Since the number of propositions in the buffer is a parameter in Kintsch and van Dijk's model, and since we were uncertain about how to pick a value for this parameter, a conservative value was adopted. We counted the number of arguments in common between the critical sentence and the previous sentences in the text; that is, the repetition count started at the critical sentence and went back to the beginning of the story. Pronouns were counted as repetitions, as were singulars and plurals, and such a change as "island" for "islanders." The number of repetitions was small, but exactly equal numbers of repetitions occurred for high-level propositions as for low-level propositions (six). Thus, there is little support in these data for the importance of this aspect of Kintsch and van Dijk's analysis. Subjects apparently did not take longer to read high-level sentences because there was less overlap between the arguments of these sentences and their context, as compared to the overlap for low-level sentences. Recall, however, that the overlap measure is only one way (though an important and explicit one) that the reading-time effect can be accounted for within the Kintsch and van Dijk model. The high-level sentences might instead have been processed longer because the macrorules applied to them or because the schema deemed them immediately relevant for storage in the buffer memory. As noted earlier, however, these latter aspects of the theory are not as well worked out as are the microprocessing parts of it. It is important to reiterate that the present materials were not designed to test the "argument overlap" hypothesis; it is fortuitous that the overlap measure between high- and low-level sentences was the same. Studies sampling a wide range of overlap values may find evidence consistent with the Kintsch and van Dijk model; but we did not find such evidence here.

It is plausible that Kintsch and van Dijk's measure of argument overlap will not be a good predictor of reading-time differences, even though it may often correlate reasonably well with some measure of text cohesiveness in samples of normal prose. Reading times may reflect difficulty of integration due to unclarities about causal connections, shifts from location to theme, and so on. One can make such transitions even when there is considerable overlap in arguments between adjacent propositions. Kintsch and van Dijk's measure is probably too syntactic in nature and not sufficiently semantic, as they themselves note. When the reader understands a text he or she constructs a set of elements and relations among them; creates, in other words, a possible world within which the events described by the discourse are occurring. It is reasonable to argue that overlap among elements in the constructed world is a better measure of text cohesion than is overlap between arguments of the propositions. The former is a semantic construct, the latter syntactic. Likewise, natural inference systems are perhaps better conceived of as working within the semantic rather than the syntactic domain. Thus, it may be that the sequence, "He could no longer talk at all. No sound came from his throat," is relatively easy to understand because the listener determines that there is no possible world in which the second proposition is true and the first false. Use of the technical tools from modern semantics (e.g., Cresswell, 1978; Lewis, 1972) is worth exploring as psychologists try to understand the understanding of texts.

The two general approaches that we have explored to account for our data are consistent with a distinction made by Frederiksen (1977) between two approaches to discourse structure and comprehension. Our first hypothesis was, in Frederiksen's terms, essentially schema based, prior knowledge being used in conjunction with cues to construct the macrostructure of the story. Our second account, however, was text based, starting from the propositions of the story itself (rather than from conventional schema) and building the structure through rules of inference differing in their ease of application. This dichotomy runs throughout much recent work in the area of text processing. It would probably be simplistic to expect that one of these general (indeed, vague) approaches to text processing was the single correct way to approach the problem of text comprehension. Further research must attempt to clarify these ideas and to narrow down the options within each approach so that data may be used to falsify classes of models devised within them.

Finally, there is the influence of serial position on reading time to be considered. Earlier we noted that reading-time differences due to serial position may arise out of differences in semantic information content across stories. De Beaugrande (1978) has argued this point. Again, much work needs to be done before this idea can be made precise within a model of text processing. For now, our results are an indication that serial position is a variable, perhaps moderated by numerous others.

Clearly, the present work only touches the surface of the problems involved in comprehending texts. But it has shown both that it is possible to tap the ongoing processes of text comprehension and that such processes are affected by the overall hierarchical organization of the text. The present work also appears to bear upon the most precise model yet proposed for text memory, that of Kintsch and van Dijk (1978). Although much remains to be done, both theoretically and empirically, we can say that our results support the fundamental intuition that underlies work in this area, namely, that there is more to understanding a text than the understanding of its sentences.

References

Bartlett, F. C. *Remembering: A study in experimental and social psychology.* London: Cambridge Univ. Press, 1932.

Bower, G. H. Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology,* 1976, **28**, 511–534.

Clark, H. H. Bridging. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science.* London: Cambridge Univ. Press, 1977.

Cresswell, M. J. Semantic competence. In F. Guenther & M. Guenther-Reutter (Eds.), *Meaning and translation.* London: Duckworth, 1978.

De Beaugrande, R. Information, expectation, and processing: On classifying poetic texts. *Poetics,* 1978, **7**, 3–44.

Fodor, J. A., Bever, T. G., & Garrett, M. G. *The psychology of language: An introduction to psycholinguistics and generative grammar.* New York: McGraw–Hill, 1974.

Frederiksen, C. H. Structure and process in discourse production and comprehension. In M. A. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Grimes, J. *The thread of discourse.* The Hague: Mouton, 1975.

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

Kintsch, W. On comprehending stories. In M.A. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension.* Hillside, N.J.: Lawrence Erlbaum, Associates, 1977.

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.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior,* 1975, **14**, 196–214.

KINTSCH, W., & VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review*, 1978, **85**, 363–394.

LEWIS, D. K. General semantics. In D. Davidson & G. H. Harman (Eds.), *Semantics of natural language*. Dordrecht: Reidel, 1972.

MANDLER, J. M. A code in the node: The use of a story schema in retrieval. *Discourse Processes*, 1978, **1**, 14–35.

MANDLER, J. M., & JOHNSON, N. J. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, **9**, 111–151.

MEYER, B. *The organization of prose and its effect upon memory*. Amsterdam: North–Holland, 1975.

OLSON, G. M., DUFFY, S. A., & MACK, R. L. Knowledge of writing conventions in prose comprehension. In W. J. McKeachie & K. Eble (Eds.), *New directions in learning and teaching*. San Francisco: Jossey–Bass, in press.

RUMELHART, D. E. Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

THORNDYKE, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, **9**, 77–110.

WINOGRAD, T. A framework for understanding discourse. In M. A. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

(Received June 25, 1979)