

Psychological Review

VOLUME 85 NUMBER 5 SEPTEMBER 1978

Toward a Model of Text Comprehension and Production

Walter Kintsch
University of Colorado

Teun A. van Dijk
University of Amsterdam,
Amsterdam, The Netherlands

The semantic structure of texts can be described both at the local microlevel and at a more global macrolevel. A model for text comprehension based on this notion accounts for the formation of a coherent semantic text base in terms of a cyclical process constrained by limitations of working memory. Furthermore, the model includes macro-operators, whose purpose is to reduce the information in a text base to its gist, that is, the theoretical macrostructure. These operations are under the control of a schema, which is a theoretical formulation of the comprehender's goals. The macroprocesses are predictable only when the control schema can be made explicit. On the production side, the model is concerned with the generation of recall and summarization protocols. This process is partly reproductive and partly constructive, involving the inverse operation of the macro-operators. The model is applied to a paragraph from a psychological research report, and methods for the empirical testing of the model are developed.

The main goal of this article is to describe the system of mental operations that underlie the processes occurring in text comprehension and in the production of recall and summarization protocols. A processing model will be outlined that specifies three sets of operations. First, the meaning elements of a text become

organized into a coherent whole, a process that results in multiple processing of some elements and, hence, in differential retention. A second set of operations condenses the full meaning of the text into its gist. These processes are complemented by a third set of operations that generate new texts from the memorial consequences of the comprehension processes.

This research was supported by Grant MH15872 from the National Institute of Mental Health to the first author and by grants from the Netherlands Organization for the Advancement of Pure Research and the Faculty of Letters of the University of Amsterdam to the second author.

We have benefited from many discussions with members of our laboratories both at Amsterdam and Colorado and especially with Ely Kozminsky, who also performed the statistical analyses reported here.

Requests for reprints should be sent to Walter Kintsch, Department of Psychology, University of Colorado, Boulder, Colorado 80309.

These goals involve a number of more concrete objectives. We want first to be able to go through a text, sentence by sentence, specifying the processes that these sentences undergo in comprehension as well as the outputs of these processes at various stages of comprehension. Next, we propose to analyze recall protocols and summaries in the same way and to specify for each sentence the operations required to produce such a sentence. The

model is, however, not merely descriptive. It will yield both simulated output protocols that can be qualitatively compared with experimental protocols and detailed predictions of the frequencies with which propositions from the text as well as certain kinds of inferences about it appear in experimental protocols, which can then be tested with conventional statistical methods.

LaBerge and Samuels (1974), in their influential article on reading processes, have remarked that "the complexity of the comprehension operation appears to be as enormous as that of thinking in general" (p. 320). As long as one accepts this viewpoint, it would be foolish to offer a model of comprehension processes. If it were not possible to separate aspects of the total comprehension process for study, our enterprise would be futile. We believe, however, that the comprehension process can be decomposed into components, some of which may be manageable at present, while others can be put aside until later.

Two very important components will be excluded from the present model. First, the model will be concerned only with semantic structures. A full grammar, including a parser, which is necessary both for the interpretation of input sentences and for the production of output sentences, will not be included. In fact, no such grammar is available now, nor is there hope for one in the near future. Hence, the model operates at the level of assumed underlying semantic structures, which we characterize in terms of propositions. Furthermore, comprehension always involves knowledge use and inference processes. The model does not specify the details of these processes. That is, the model only says when an inference occurs and what it will be; the model does not say how it is arrived at, nor what precisely was its knowledge base. Again, this restriction is necessary because a general theory of inference processes is nowhere in sight, and the alternative of restricting the model to a small but manageable base has many disadvantages.

Comprehension can be modeled only if we are given a specific goal. In other words, we must know the control processes involved in comprehension. Frequently, these are ill specified, and thus we must select cases where well-defined control processes do exist, prefer-

ably ones that are shared by many readers in order to make feasible the collection of meaningful data. Hence, we shall emphasize the processing of certain conventional text types, such as stories.

Comprehension is involved in reading as well as in listening, and our model applies to both. Indeed, the main differences between reading and listening occur at levels lower than the ones we are concerned with (e.g., Kintsch & Kozminsky, 1977; Sticht, in press). We are talking here, of course, about readers for whom the decoding process has become automated, but the model also has implications for readers who still must devote substantial resources to decoding.

One of the characteristics of the present model is that it assumes a multiplicity of processes occurring sometimes in parallel, sometimes sequentially. The fact that, phenomenologically, comprehension is a very simple experience does not disprove this notion of multiple, overlapping processes. Perception is likewise a unitary, direct experience; but when its mechanisms are analyzed at the physiological and psychological levels, they are found to consist of numerous interacting subprocesses that run off rapidly and with very little mutual interference. Resources seem to be required only as attention, consciousness, decisions, and memory become involved; it is here that the well-known capacity limitations of the human system seem to be located rather than in the actual processing. Thus, in the comprehension model presented below, we assume several complex processes operating in parallel and interactively without straining the resources of the system. Capacity limitations become crucial, however, when it comes to the storage of information in memory and response production. That such a system is very complex should not in itself make it implausible as an account of comprehension processes.

The organization of this article is as follows. First, the semantic structure of texts, including their macrostructure, is discussed briefly, mainly with reference to our previous work. Then, a psychological processing model is described that is based on these ideas about text structure. While related, the functions of these two enterprises are rather different. The

structural theory is a semiformal statement of certain linguistic intuitions, while the processing model attempts to predict the contents of experimental protocols. Therefore, they must be evaluated in rather different ways. Finally, some preliminary applications and tests of the processing model are described, and some linguistic considerations are raised that may guide future extensions of it.

Semantic Structures

We assume that the surface structure of a discourse is interpreted as a set of propositions. This set is ordered by various semantic relations among the propositions. Some of these relations are explicitly expressed in the surface structure of the discourse; others are inferred during the process of interpretation with the help of various kinds of context-specific or general knowledge.

The semantic structure of a discourse is characterized at two levels, namely, at the levels of *microstructure* and of *macrostructure*. The microstructure is the local level of the discourse, that is, the structure of the individual propositions and their relations. The macrostructure is of a more global nature, characterizing the discourse as a whole. These levels are related by a set of specific semantic mapping rules, the *macrorules*. At both levels, we provide an account of the intuitive notion of the coherence of a discourse: A discourse is coherent only if its respective sentences and propositions are connected, and if these propositions are organized globally at the macrostructure level.

Microstructure of Discourse

In our earlier work (Kintsch, 1974; van Dijk, 1972, 1977d), we attempted to characterize the semantic structure of a discourse in terms of an abstract *text base*. The present article will shift attention to the more specific properties of comprehension, that is, to the ways a language understander actually constructs such a text base.

Text bases are not merely unrelated lists of propositions; they are coherent, structured units. One way to assign a structure to a text base may be derived from its referential

coherence: We establish a linear or hierarchical sequence of propositions in which coreferential expressions occur. The first (or superordinate) of these propositions often appears to have a specific cognitive status in such a sequence, being recalled two or three times more often than other propositions (see Kintsch, 1974). Clearly, a processing model must account for the superior recall of propositions that function as superordinates in the text-base structure.

Inferences

Natural language discourse may be connected even if the propositions expressed by the discourse are not directly connected. This possibility is due to the fact that language users are able to provide, during comprehension, the missing links of a sequence on the basis of their general or contextual knowledge of the facts. In other words, the facts, as known, allow them to make inferences about possible, likely, or necessary other facts and to interpolate missing propositions that may make the sequence coherent. Given the general pragmatic postulate that we do not normally assert what we assume to be known by the listener, a speaker may leave all propositions implicit that can be provided by the listener. An actual discourse, therefore, normally expresses what may be called an *implicit* text base. An *explicit* text base, then, is a theoretical construct featuring also those propositions necessary to establish formal coherence. This means that only those propositions are interpolated that are interpretation conditions, for example, presuppositions of other propositions expressed by the discourse.

Macrostructure of Discourse

The semantic structure of discourse must be described not only at the local microlevel but also at the more global macrolevel. Besides the purely psychological motivation for this approach, which will become apparent below in our description of the processing model, the theoretical and linguistic reasons for this level of description derive from the fact that the propositions of a text base must be connected relative to what is intuitively called a

topic of discourse (or a topic of conversation), that is, the theme of the discourse or a fragment thereof. Relating propositions in a local manner is not sufficient. There must be a global constraint that establishes a meaningful whole, characterized in terms of a discourse topic.

The notion of a discourse topic can be made explicit in terms of semantic macrostructures. Like other semantic structures, these macrostructures are described in terms of propositions and proposition sequences. In order to show how a discourse topic is related to the respective propositions of the text base, we thus need semantic mapping rules with microstructural information as input and macrostructural information as output. Such macrorules, as we call them, both reduce and organize the more detailed information of the microstructure of the text. They describe the same facts but from a more global point of view. This means, for instance, that we may have several levels of macrostructure; macrorules may apply recursively on sequences of macropropositions as long as the constraints on the rules are satisfied.

The general abstract nature of the macrorules is based on the relation of semantic entailment. That is, they preserve both truth and meaning: A macrostructure must be implied by the (explicit) microstructure from which it is derived. Note that in actual processing, as we will see, macrostructures may also be inferred on the basis of incomplete information.

The basic constraint of the macrorules is that no proposition may be deleted that is an interpretation condition of a following proposition of the text. In fact, this also guarantees that a macrostructure itself is connected and coherent.

A formal description of the macrorules will not be given here. We shall merely mention them briefly in an informal way (for details, see van Dijk, 1977b, 1977d).

1. *Deletion.* Each proposition that is neither a direct nor an indirect interpretation condition of a subsequent proposition may be deleted.

2. *Generalization.* Each sequence of propositions may be substituted by the general proposition denoting an immediate superset.

3. *Construction.* Each sequence of proposi-

tions may be substituted by a proposition denoting a global fact of which the facts denoted by the microstructure propositions are normal conditions, components, or consequences.

The macrorules are applied under the control of a schema, which constrains their operation so that macrostructures do not become virtually meaningless abstractions or generalizations.

Just as general information is needed to establish connection and coherence at the microstructural level, world knowledge is also required for the operation of macrorules. This is particularly obvious in Rule 3, where our frame knowledge (Minsky, 1975) specifies which facts belong conventionally to a more global fact, for example, "paying" as a normal component of both "shopping" and "eating in a restaurant" (cf. Bobrow & Collins, 1975).

Schematic Structures of Discourse

Typical examples of conventional schematic structures of discourse are the structure of a story, the structure of an argument, or the structure of a psychological report. These structures are specified by a set of characteristic categories and a set of (sometimes recursive) rules of formation and transformation defining the canonical and possible ordering of the categories (e.g., van Dijk, 1976, *in press*).

Schematic structures play an important role in discourse comprehension and production. Without them, we would not be able to explain why language users are able to understand a discourse as a story, or why they are able to judge whether a story or an argument is correct or not. More specifically, we must be able to explain why a free-recall protocol or a summary of a story also has a narrative structure, or why the global structure of a psychological paper is often conserved in its abstract (cf. Kintsch & van Dijk, 1975; van Dijk & Kintsch, 1977). Furthermore, schematic structures play an important control function in the processing model to be described below, where they are used not only in connection with conventional texts but also to represent idiosyncratic personal processing goals. The application of the macrorules depends on whether a given proposition is or is not judged

to be relevant in its context, with the schema specifying the kind of information that is to be considered relevant for a particular comprehension task.

Process Model

Forming Coherent Text Bases

The model takes as its input a list of propositions that represent the meaning of a text. The derivation of this proposition list from the text is not part of the model. It has been described and justified in Kintsch (1974) and further codified by Turner and Greene (in press). With some practice, persons using this system arrive at essentially identical propositional representations, in our experience. Thus, the system is simple to use; and while it is undoubtedly too simple to capture nuances of meaning that are crucial for some linguistic and logical analyses, its robustness is well suited for a process model. Most important, this notation can be translated into other systems quite readily. For instance, one could translate the present text bases into the graphical notation of Norman and Rumelhart (1975), though the result would be quite cumbersome, or into a somewhat more sophisticated notation modeled after the predicate calculus, which employs atomic propositions, variables, and constants (as used in van Dijk, 1973). Indeed, the latter notation may eventually prove to be more suitable. The important point to note is simply that although some adequate notational system for the representation of meaning is required, the details of that system often are not crucial for our model.

Propositional Notation

The propositional notation employed below will be outlined here only in the briefest possible way. The idea is to represent the meaning of a text by means of a structured list of *propositions*. Propositions are composed of *concepts* (the names of which we shall write in capital letters, so that they will not be confused with words). The composition rule states that each proposition must include first a predicate, or *relational concept*, and one or more *arguments*.

The latter may be concepts or other embedded propositions. The arguments of a proposition fulfill different semantic functions, such as agent, object, and goal. Predicates may be realized in the surface structure as verbs, adjectives, adverbs, and sentence connectives. Each predicate constrains the nature of the argument that it may take. These constraints are imposed both by linguistic rules and general world knowledge and are assumed to be a part of a person's knowledge or semantic memory.

Propositions are ordered in the text base according to the way in which they are expressed in the text itself. Specifically, their order is determined by the order of the words in the text that correspond to the propositional predicates.

Text bases must be coherent. One of the linguistic criteria for the semantic coherence of a text base is referential coherence. In terms of the present notational system, referential coherence corresponds to argument overlap among propositions. Specifically, (P, A, B) is referentially coherent with (R, B, C) because the two propositions share the argument B, or with (Q, D, (P, A, B)) because one proposition is embedded here as an argument into another. Referential coherence is probably the most important single criterion for the coherence of text bases. It is neither a necessary nor a sufficient criterion linguistically. However, the fact that in many texts other factors tend to be correlated with it makes it a useful indicator of coherence that can be checked easily, quickly, and reliably. We therefore propose the following hypothesis about text processing: The first step in forming a coherent text base consists in checking out its referential coherence; if a text base is found to be referentially coherent, that is, if there is some argument overlap among all of its propositions, it is accepted for further processing; if gaps are found, inference processes are initiated to close them; specifically, one or more propositions will be added to the text base that make it coherent. Note that in both cases, the argument-referent repetition rule also holds for arguments consisting of a proposition, thereby establishing relations not only between individuals but also between facts denoted by propositions.

Processing Cycles

The second major assumption is that this checking of the text base for referential coherence and the addition of inferences wherever necessary cannot be performed on the text base as a whole because of the capacity limitations of working memory. We assume that a text is processed sequentially from left to right (or, for auditory inputs, in temporal order) in chunks of several propositions at a time. Since the proposition lists of a text base are ordered according to their appearance in the text, this means that the first n_1 propositions are processed together in one cycle, then the next n_2 propositions, and so on. It is unreasonable to assume that all n_i s are equal. Instead, for a given text and a given comprehender, the maximum n_i will be specified. Within this limitation, the precise number of propositions included in a processing chunk depends on the surface characteristics of the text. There is ample evidence (e.g., Aaronson & Scarborough, 1977; Jarvella, 1971) that sentence and phrase boundaries determine the chunking of a text in short-term memory. The maximum value of n_i , on the other hand, is a model parameter, depending on text as well as reader characteristics.¹

If text bases are processed in cycles, it becomes necessary to make some provision in the model for connecting each chunk to the ones already processed. The following assumptions are made. Part of working memory is a short-term memory buffer of limited size s . When a chunk of n_i propositions is processed, s of them are selected and stored in the buffer.² Only those s propositions retained in the buffer are available for connecting the new incoming chunk with the already processed material. If a connection is found between any of the new propositions and those retained in the buffer, that is, if there exists some argument overlap between the input set and the contents of the short-term memory buffer, the input is accepted as coherent with the previous text. If not, a resource-consuming search of all previously processed propositions is made. In auditory comprehension, this search ranges over all text propositions stored in long-term memory. (The storage assumptions are detailed below.) In reading comprehension, the search includes all previous propositions because even

those propositions not available from long-term memory can be located by re-reading the text itself. Clearly, we are assuming here a reader who processes all available information; deviant cases will be discussed later.

If this search process is successful, that is, if a proposition is found that shares an argument with at least one proposition in the input set, the set is accepted and processing continues. If not, an inference process is initiated, which adds to the text base one or more propositions that connect the input set to the already processed propositions. Again, we are assuming here a comprehender who fully processes the text. Inferences, like long-term memory searches, are assumed to make relatively heavy demands on the comprehender's resources and, hence, contribute significantly to the difficulty of comprehension.

Coherence Graphs

In this manner, the model proceeds through the whole text, constructing a network of coherent propositions. It is often useful to represent this network as a graph, the nodes of which are propositions and the connecting lines indicating shared referents. The graph can be arranged in levels by selecting that proposition for the top level that results in the simplest graph structure (in terms of some suitable measure of graph complexity). The second level in the graph is then formed by all the propositions connected to the top level; propositions that are connected to any proposi-

¹ Presumably, a speaker employs surface cues to signal to the comprehender an appropriate chunk size. Thus, a good speaker attempts to place his or her sentence boundaries in such a way that the listener can use them effectively for chunking purposes. If a speaker-listener mismatch occurs in this respect, comprehension problems arise because the listener cannot use the most obvious cues provided and must rely on harder-to-use secondary indicators of chunk boundaries. One may speculate that sentence grammars evolved because of the capacity limitations of the human cognitive system, that is, from a need to package information in chunks of suitable size for a cyclical comprehension process.

² Sometimes a speaker does not entrust the listener with this selection process and begins each new sentence by repeating a crucial phrase from the old one. Indeed, in some languages such "backreference" is obligatory (Longacre & Levinsohn, 1977).

tion at the second level, but not to the first-level proposition, then form the third level. Lower levels may be constructed similarly. For convenience, in drawing such graphs, only connections between levels, not within levels, are indicated. In case of multiple between-level connections, we arbitrarily indicate only one (the first one established in the processing order). Thus, each set of propositions, depending on its pattern of interconnections, can be assigned a unique graph structure.

The topmost propositions in this kind of coherence graph may represent presuppositions of their subordinate propositions due to the fact that they introduce relevant discourse referents, where presuppositions are important for the contextual interpretation and hence for discourse coherence, and where the number of subordinates may point to a macrostructural role of such presupposed propositions. Note that the procedure is strictly formal: There is no claim that topmost propositions are always most important or relevant in a more intuitive sense. That will be taken care of with the macro-operations described below.

A coherent text base is therefore a connected graph. Of course, if the long-term memory searches and inference processes required by the model are not executed, the resulting text base will be incoherent, that is, the graph will consist of several separate clusters. In the experimental situations we are concerned with, where the subjects read the text rather carefully, it is usually reasonable to assume that coherent text bases are established.

To review, so far we have assumed a cyclical process that checks on the argument overlap in the proposition list. This process is automatic, that is, it has low resource requirements. In each cycle, certain propositions are retained in a short-term buffer to be connected with the input set of the next cycle. If no connections are found, resource-consuming search and inference operations are required.

Memory Storage

In each processing cycle, there are n_i propositions involved, plus the s propositions held over in the short-term buffer. It is assumed that in each cycle, the propositions

currently being processed may be stored in long-term memory and later reproduced in a recall or summarization task, each with probability p . This probability p is called the *reproduction probability* because it combines both storage and retrieval information. Thus, for the same comprehension conditions, the value of p may vary depending on whether the subject's task is recall or summarization or on the length of the retention interval. We are merely saying that for some comprehension-production combination, a proposition is reproduced with probability p for each time it has participated in a processing cycle. Since at each cycle a subset of s propositions is selected and held over to the next processing cycle, some propositions will participate in more than one processing cycle and, hence, will have higher reproduction probabilities. Specifically, if a proposition is selected $k - 1$ times for inclusion in the short-term memory buffer, it has k chances of being stored in long-term memory, and hence, its reproduction probability will be $1 - (1 - p)^k$.

Which propositions of a text base will be thus favored by multiple processing depends crucially on the nature of the process that selects the propositions to be held over from one processing cycle to the next. Various strategies may be employed. For instance, the s buffer propositions may be chosen at random. Clearly, such a strategy would be less than optimal and result in unnecessary long-term memory searches and, hence, in poor reading performance. It is not possible at this time to specify a unique optimal strategy. However, there are two considerations that probably characterize good strategies. First, a good strategy should select propositions for the buffer that are important, in the sense that they play an important role in the graph already constructed. A proposition that is already connected to many other propositions is more likely to be relevant to the next input cycle than a proposition that has played a less pivotal role so far. Thus, propositions at the top levels of the graph to which many others are connected should be selected preferentially. Another reasonable consideration involves recency. If one must choose between two equally important propositions in the sense outlined above, the more recent one might be

expected to be the one most relevant to the next input cycle. Unfortunately, these considerations do not specify a unique selection strategy but a whole family of such strategies. In the example discussed in the next section, a particular example of such a strategy will be explored in detail, not because we have any reason to believe that it is better than some alternatives, but to show that each strategy leads to empirically testable consequences. Hence, which strategy is predominantly used in a given population of subjects becomes an empirically decidable issue because each strategy specifies a somewhat different pattern of reproduction probabilities over the propositions of a text base. Experimental results may be such that only one particular selection strategy can account for them, or perhaps we shall not be able to discriminate among a class of reasonable selection strategies on the basis of empirical frequency distributions. In either case, we shall have the information the model requires.

It is already clear that a random selection strategy will not account for paragraph recall data, and that some selection strategy incorporating an "importance" principle is required. The evidence comes from work on what has been termed the "levels effect" (Kintsch & Keenan, 1973; Kintsch et al., 1975; Meyer, 1975). It has been observed that propositions belonging to high levels of a text-base hierarchy are much better recalled (by a factor of two or three) than propositions low in the hierarchy. The text-base hierarchies in these studies were constructed as follows. The topical proposition or propositions of a paragraph (as determined by its title or otherwise by intuition) were selected as the top level of the hierarchy, and all those propositions connected to them via argument overlap were determined, forming the next level of the hierarchy, and so on. Thus, these hierarchies were based on referential coherence, that is, arguments overlap among propositions. In fact, the propositional networks constructed here are identical with these hierarchies, except for changes introduced by the fact that in the present model, coherence graphs are constructed cycle by cycle.

Note that the present model suggests an interesting reinterpretation of this levels effect.

Suppose that the selection strategy, as discussed above, is indeed biased in favor of important, that is, high-level propositions. Then, such propositions will, on the average, be processed more frequently than low-level propositions and recalled better. The better recall of high-level propositions can then be explained because they are processed differently than low-level propositions. Thus, we are now able to add a processing explanation to our previous account of levels effects, which was merely in terms of a correlation between some structural aspects of texts and recall probabilities.

Although we have been concerned here only with the organization of the micropropositions, one should not forget that other processes, such as macro-operations, are going on at the same time, and that therefore the buffer must contain the information about macropropositions and presuppositions that is required to establish the global coherence of the discourse. One could extend the model in that direction. Instead, it appears to us worthwhile to study these various processing components in isolation, in order to reduce the complexity of our problem to manageable proportions. We shall remark on the feasibility of this research strategy below.

Testing the Model

The frequencies with which different text propositions are recalled provide the major possibility for testing the model experimentally. It is possible to obtain good estimates of such frequencies for a given population of readers and texts. Then, the best-fitting model can be determined by varying the three parameters of the model— n , the maximum input size per cycle; s , the capacity of the short-term buffer; and p , the reproduction probability—at the same time exploring different selection strategies. Thus, which selection strategy is used by a given population of readers and for a given type of text becomes decidable empirically. For instance, in the example described below, a selection strategy has been assumed that emphasizes recency and frequency (the "leading-edge" strategy). It is conceivable that for college students reading texts rather carefully under laboratory conditions, a fairly sophisticated

strategy like this might actually be used (i.e., it would lead to fits of the model detectably better than alternative assumptions). At the same time, different selection strategies might be used in other reader populations or under different reading conditions. It would be worthwhile to investigate whether there are strategy differences between good and poor readers, for instance.

If the model is successful, the values of the parameters s , n , and p , and their dependence on text and reader characteristics might also prove to be informative. A number of factors might influence s , the short-term memory capacity. Individuals differ in their short-term memory capacity. Perfetti and Goldman (1976), for example, have shown that good readers are capable of holding more of a text in short-term memory than poor readers. At the same time, good and poor readers did not differ on a conventional memory-span test. Thus, the differences observed in the reading task may not be due to capacity differences per se. Instead, they could be related to the observation of Hunt, Lunneborg, and Lewis (1975) that persons with low verbal abilities are slower in accessing information in short-term memory. According to the model, a comprehender continually tests input propositions against the contents of the short-term buffer; even a slight decrease in the speed with which these individual operations are performed would result in a noticeable deterioration in performance. In effect, lowering the speed of scanning and matching operations would have the same effect as decreasing the capacity of the buffer.

The buffer capacity may also depend on the difficulty of the text, however, or more precisely, on how difficult particular readers find a text. Presumably, the size of the buffer depends, within some limits, on the amount of resources that must be devoted to other aspects of processing (perceptual decoding, syntactic-semantic analyses, inference generation, and the macro-operations discussed below). The greater the automaticity of these processes and the fewer inferences required, the larger the buffer a reader has to operate with. Certainly, familiarity should have a pronounced effect on n , the number of propositions accepted per cycle. The process of construct-

ing a text base is perhaps best described as *apperception*. That is, a reader's knowledge determines to a large extent the meaning that he or she derives from a text. If the knowledge base is lacking, the reader will not be able to derive the same meaning that a person with adequate knowledge, reading the same text, would obtain. Unfamiliar material would have to be processed in smaller chunks than familiar material, and hence, n should be directly related to familiarity. No experimental tests of this prediction appear to exist at present. However, other factors, too, might influence n . Since it is determined by the surface form of the text, increasing the complexity of the surface form while leaving the underlying meaning intact ought to decrease the size of the processing chunks. Again, no data are presently available, though Kintsch and Monk (1972) and King and Greeno (1974) have reported that such manipulations affect reading times.

Finally, the third parameter of the model, p , would also be expected to depend on familiarity for much the same reasons as s : The more familiar a text, the fewer resources are required for other aspects of processing, and the more resources are available to store individual propositions in memory. The value of p should, however, depend above all on the task demands that govern the comprehension process as well as the later production process. If a long text is read with attention focused mainly on gist comprehension, the probability of storing individual propositions of the text base should be considerably lower than when a short paragraph is read with immediate recall instructions. Similarly, when summarizing a text, the value of p should be lower than in recall because individual micropropositions are given less weight in a summary relative to macropropositions.

Note that in a model like the present one, such factors as familiarity may have rather complex effects. Not only can familiar material be processed in larger chunks and retained more efficiently in the memory buffer; if a topic is unfamiliar, there will be no frame available to organize and interpret a given proposition sequence (e.g., for the purpose of generating inferences from it), so readers might continue to pick up new propositions in the hope of finding information that may organize what

they already hold in their buffer. If, however, the crucial piece of information fails to arrive, the working memory will be quickly overloaded and incomprehension will result. Thus, familiarity may have effects on comprehension not only at the level of processing considered here (the construction of a coherent text base) but also at higher processing levels.

Readability

The model's predictions are relevant not only for recall but also for the readability of texts. This aspect of the model has been explored by Kintsch and Vipond (1978) and will be only briefly summarized here. It was noted there that conventional accounts of readability have certain shortcomings, in part because they do not concern themselves with the organization of long texts, which might be expected to be important for the ease of difficulty with which a text can be comprehended. The present model makes some predictions about readability that go beyond these traditional accounts. Comprehension in the normal case is a fully automatic process, that is, it makes low demands on resources. Sometimes, however, these normal processes are blocked, for instance, when a reader must retrieve a referent no longer available in his or her working memory, which is done almost consciously, requiring considerable processing resources. One can simulate the comprehension process according to the model and determine the number of resource-consuming operations in this process, that is, the number of long-term memory searches required and the number of inferences required. The assumption was made that each one of these operations disrupts the automatic comprehension processes and adds to the difficulty of reading. On the other hand, if these operations are not performed by a reader, the representation of the text that this reader arrives at will be incoherent. Hence, the retrieval of the text base and all performance depending on its retrieval (such as recall, summarizing, or question answering) will be poor. Texts requiring many operations that make high demands on resources should yield either increased reading times or low scores on comprehension tests. Comprehension, therefore,

must be evaluated in a way that considers both comprehension time and test performance because of the trade-off between the two. Only measures such as reading time per proposition recalled (used in Kintsch et al., 1975) or reading speed adjusted for comprehension (suggested by Jackson & McClelland, 1975) can therefore be considered adequate measures of comprehension difficulty.

The factors related to readability in the model depend, of course, on the input size per cycle (n) and the short-term memory capacity (s) that are assumed. In addition, they depend on the nature of the selection strategy used, since that determines which propositions in each cycle are retained in the short-term buffer to be interconnected with the next set of input propositions. A reader with a poor selection strategy and a small buffer, reading unfamiliar material, might have all kinds of problems with a text that would be highly readable for a good reader. Thus, readability cannot be considered a property of texts alone, but one of the text-reader interaction. Indeed, some preliminary analyses reported by Kintsch and Vipond (1978) show that the readability of some texts changes a great deal as a function of the short-term memory capacity and the size of input chunks: Some texts were hard for all parameter combinations that were explored; others were easy in every case; still others could be processed by the model easily when short-term memory and input chunks were large, yet became very difficult for small values of these parameters.

Macrostructure of a Text

Macro-operators transform the propositions of a text base into a set of macropropositions that represent the gist of the text. They do so by deleting or generalizing all propositions that are either irrelevant or redundant and by constructing new inferred propositions. "Delete" here does not mean "delete from memory" but "delete from the macrostructure." Thus, a given text proposition—a microproposition—may be deleted from the text's macrostructure but, nevertheless, be stored in memory and subsequently recalled as a microproposition.

Role of the Schema

The reader's goals in reading control the application of the macro-operators. The formal representation of these goals is the *schema*. The schema determines which micropropositions or generalizations of micropropositions are relevant and, thus, which parts of the text will form its gist.

It is assumed that text comprehension is always controlled by a specific schema. However, in some situations, the controlling schema may not be detailed, nor predictable. If a reader's goals are vague, and the text that he or she reads lacks a conventional structure, different schemata might be set up by different readers, essentially in an unpredictable manner. In such cases, the macro-operations would also be unpredictable. Research on comprehension must concentrate on those cases where texts are read with clear goals that are shared among readers. Two kinds of situations qualify in this respect. First of all, there are a number of highly conventionalized text types. If a reader processes such texts in accordance with their conventional nature, specific well-defined schemata are obtained. These are shared by the members of a given cultural group and, hence, are highly suitable for research purposes. Familiar examples of such texts are stories (e.g., Kintsch & van Dijk, 1975) and psychological research reports (Kintsch, 1974). These schemata specify both the schematic categories of the texts (e.g., a research report is supposed to contain introduction, method, results, and discussion sections), as well as what information in each section is relevant to the macrostructure (e.g., the introduction of a research report must specify the purpose of the study). Predictions about the macrostructure of such texts require a thorough understanding of the nature of their schemata. Following the lead of anthropologists (e.g., Colby, 1973) and linguists (e.g., Labov & Waletzky, 1967), psychologists have so far given the most attention to story schemata, in part within the present framework (Kintsch, 1977; Kintsch & Greene, 1978; Poulsen, Kintsch, Kintsch, & Premack, in press; van Dijk, 1977b) and in part within the related "story grammar" approach (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, in press).

A second type of reading situation where well-defined schemata exist comprises those cases where one reads with a special purpose in mind. For instance, Hayes, Waterman, and Robinson (1977) have studied the relevance judgments of subjects who read a text with a specific problem-solving set. It is not necessary that the text be conventionally structured. Indeed, the special purpose overrides whatever text structure there is. For instance, one may read a story with the processing controlled not by the usual story schema but by some special-purpose schema established by task instructions, special interests, or the like. Thus, *Decameron* stories may be read not for the plot and the interesting events but because of concern with the role of women in fourteenth-century Italy or with the attitudes of the characters in the story toward morality and sin.

There are many situations where comprehension processes are similarly controlled. As one final example, consider the case of reading the review of a play with the purpose of deciding whether or not to go to that play. What is important in a play for the particular reader serves as a schema controlling the gist formation: There are open slots in this schema, each one standing for a property of plays that this reader finds important, positively or negatively, and the reader will try to fill these slots from the information provided in the review. Certain propositions in the text base of the review will be marked as relevant and assigned to one of these slots, while others will be disregarded. If thereby not all slots are filled, inference processes will take over, and an attempt will be made to fill in the missing information by applying available knowledge frames to the information presented directly. Thus, while the macro-operations themselves are always information reducing, the macrostructure may also contain information not directly represented in the original text base, when such information is required by the controlling schema. If the introduction to a research report does not contain a statement of the study's purpose, one will be inferred according to the present model.

In many cases, of course, people read loosely structured texts with no clear goals in mind. The outcome of such comprehension

processes, as far as the resulting macrostructure is concerned, is indeterminate. We believe, however, that this is not a failure of the model: If the schema that controls the macro-operations is not well defined, the outcome will be haphazard, and we would argue that no scientific theory, in principle, can predict it.

Macro-operators

The schema thus classifies all propositions of a text base as either relevant or irrelevant. Some of these propositions have generalizations or constructions that are also classified in the same way. In the absence of a general theory of knowledge, whether a microproposition is generalizable in its particular context or can be replaced by a construction must be decided on the basis of intuition at present. Each microproposition may thus be either deleted from the macrostructure or included in the macrostructure as is, or its generalization or construction may be included in the macrostructure. A proposition or its generalization/construction that is included in the macrostructure is called a *macroproposition*. Thus, some propositions may be both micro- and macropropositions (when they are relevant); irrelevant propositions are only micropropositions, and generalizations and constructions are only macropropositions.³

The macro-operations are performed probabilistically. Specifically, irrelevant micropropositions never become macropropositions. But, if such propositions have generalizations or constructions, their generalizations or constructions may become macropropositions, with probability g when they, too, are irrelevant, and with probability m when they are relevant. Relevant micropropositions become macropropositions with probability m ; if they have generalizations or constructions, these are included in the macrostructure, also with probability m . The parameters m and g are reproduction probabilities, just as p in the previous section: They depend on both storage and retrieval conditions. Thus, m may be small because the reading conditions did not encourage macroprocessing (e.g., only a short paragraph was read, with the expectation of an immediate recall test) or because the production was delayed for such a long time that

even the macropropositions have been forgotten.

Macrostructures are hierarchical, and hence, macro-operations are applied in several cycles, with more and more stringent criteria of relevance. At the lowest level of the macrostructure, relatively many propositions are selected as relevant by the controlling schema and the macro-operations are performed accordingly. At the next level, stricter criteria for relevance are assumed, so that only some of the first-level macropropositions are retained as second-level macropropositions. At subsequent levels, the criterion is strengthened further until only a single macroproposition (essentially a title for that text unit) remains. A macroproposition that was selected as relevant at k levels of the hierarchy has therefore a reproduction probability of $1 - (1 - m)^k$, which is the probability that at least one of the k times that this proposition was processed results in a successful reproduction.

Production

Recall or summarization protocols obtained in experiments are texts in their own right, satisfying the general textual and contextual conditions of production and communication. A protocol is not simply a replica of a memory representation of the original discourse. On the contrary, the subject will try to produce a new text that satisfies the pragmatic conditions of a particular task context in an experiment or the requirements of effective communication in a more natural context. Thus, the language user will not produce information that he or she assumes is already known or redundant. Furthermore, the operations involved in discourse production are so complex that the subject will be unable to retrieve at any one time all the information that is in principle accessible to memory. Finally, protocols will contain information not based on what the subject remembers from the original text, but

³ Longacre and Levinsohn (1977) mention a language that apparently uses macrostructure markers in the surface structure: In Cubeo, a South American Indian language, certain sentences contain a particular particle; stringing together the sentences of a text that are marked by this particle results in an abstract or summary of the text.

consisting of reconstructively added details, explanations, and various features that are the result of output constraints characterizing production in general.

Optional transformations. Propositions stored in memory may be transformed in various essentially arbitrary and unpredictable ways, in addition to the predictable, schema-controlled transformations achieved through the macro-operations that are the focus of the present model. An underlying conception may be realized in the surface structure in a variety of ways, depending on pragmatic and stylistic concerns. Reproduction (or reconstruction) of a text is possible in terms of different lexical items or larger units of meaning. Transformations may be applied at the level of the microstructure, the macrostructure, or the schematic structure. Among these transformations one can distinguish reordering, explication of coherence relations among propositions, lexical substitutions, and perspective changes. These transformations may be a source of errors in protocols, too, though most of the time they preserve meaning. Whether such transformations are made at the time of comprehension, or at the time of production, or both cannot be decided at present.

A systematic account of optional transformations might be possible, but it is not necessary within the present model. It would make an already complex model and scoring system even more cumbersome, and therefore, we choose to ignore optional transformations in the applications reported below.

Reproduction. Reproduction is the simplest operation involved in text production. A subject's memory for a particular text is a memory episode containing the following types of memory traces: (a) traces from the various perceptual and linguistic processes involved in text processing, (b) traces from the comprehension processes, and (c) contextual traces. Among the first would be memory for the typeface used or memory for particular words and phrases. The third kind of trace permits the subject to remember the circumstances in which the processing took place: the laboratory setting, his or her own reactions to the whole procedure, and so on. The model is not concerned with either of these two types of memory traces. The traces resulting from the

comprehension processes, however, are specified by the model in detail. Specifically, those traces consist of a set of micropropositions. For each microproposition in the text base, the model specifies a probability that it will be included in this memory set, depending on the number of processing cycles in which it participated and on the reproduction parameter p . In addition, the memory episode contains a set of macropropositions, with the probability that a macroproposition is included in that set being specified by the parameter m . (Note that the parameters m and p can only be specified as a joint function of storage and retrieval conditions; hence, the memory content on which a production is based must also be specified with respect to the task demands of a particular production situation.)

A reproduction operator is assumed that retrieves the memory contents as described above, so that they become part of the subject's text base from which the output protocol is derived.

Reconstruction. When micro- or macro-information is no longer directly retrievable, the language user will usually try to reconstruct this information by applying rules of inference to the information that is still available. This process is modeled with three reconstruction operators. They consist of the inverse application of the macro-operators and result in the reconstruction of some of the information deleted from the macrostructure with the aid of world knowledge: (a) addition of plausible details and normal properties, (b) particularization, and (c) specification of normal conditions, components, or consequences of events. In all cases, errors may be made. The language user may make guesses that are plausible but wrong or even give details that are inconsistent with the original text. However, the reconstruction operators are not applied blindly. They operate under the control of the schema, just as in the case of the macro-operators. Only reconstructions that are relevant in terms of this control schema are actually produced. Thus, for instance, when "Peter went to Paris by train" is a remembered macroproposition, it might be expanded by means of reconstruction operations to include such a normal component as "He went into the station to buy a ticket,"

but it would not include irrelevant reconstructions such as "His leg muscles contracted." The schema in control of the output operation need not be the same as the one that controlled comprehension: It is perfectly possible to look at a house from the standpoint of a buyer and later to change one's viewpoint to that of a prospective burglar, though different information will be produced than in the case when no such schema change has occurred (Anderson & Pichert, 1978).

Metastatements. In producing an output protocol, a subject will not only operate directly on available information but will also make all kinds of metacomments on the structure, the content, or the schema of the text. The subject may also add comments, opinions, or express his or her attitude.

Production plans. In order to monitor the production of propositions and especially the connections and coherence relations, it must be assumed that the speaker uses the available macropropositions of each fragment of the text as a production plan. For each macroproposition, the speaker reproduces or reconstructs the propositions dominated by it. Similarly, the schema, once actualized, will guide the global ordering of the production process, for example, the order in which the macropropositions themselves are actualized. At the microlevel, coherence relations will determine the ordering of the propositions to be expressed, as well as the topic-comment structure of the respective sentences. Both at the micro- and macrolevels, production is guided not only by the memory structure of the discourse itself but also by general knowledge about the normal ordering of events and episodes, general principles of ordering of events and episodes, general principles of ordering of information in discourse, and schematic rules. This explains, for instance, why speakers will often transform a non-canonical ordering into a more canonical one (e.g., when summarizing scrambled stories, as in Kintsch, Mandel, & Kozminsky, 1977).

Just as the model of the comprehension process begins at the propositional level rather than at the level of the text itself, the production model will also leave off at the propositional level. The question of how the text base that underlies a subject's protocol is

transformed into an actual text will not be considered here, although it is probably a less intractable problem than that posed by the inverse transformation of verbal texts into conceptual bases.

Text generation. Although we are concerned here with the production of second-order discourses, that is, discourses with respect to another discourse, such as free-recall protocols or summaries, the present model may eventually be incorporated into a more general theory of text production. In that case, the core propositions of a text would not be a set of propositions remembered from some specific input text, and therefore new mechanisms must be described that would generate core propositions de novo.

Processing Simulation

How the model outlined in the previous section works is illustrated here with a simple example. A suitable text for this purpose must be (a) sufficiently long to ensure the involvement of macroprocesses in comprehension; (b) well structured in terms of a schema, so that predictions from the present model become possible; and (c) understandable without technical knowledge. A 1,300-word research report by Heussenstam (1971) called "Bumperstickers and the Cops," appeared to fill these requirements. Its structure follows the conventions of psychological research reports closely, though it is written informally, without the usual subheadings indicating methods, results, and so on. Furthermore, it is concerned with a social psychology experiment that requires no previous familiarity with either psychology, experimental design, or statistics.

We cannot analyze the whole report here, though a tentative macroanalysis of "Bumperstickers" has been given in van Dijk (in press). Instead, we shall concentrate only on its first paragraph:

A series of violent, bloody encounters between police and Black Panther Party members punctuated the early summer days of 1969. Soon after, a group of Black students I teach at California State College, Los Angeles, who were members of the Panther Party, began to complain of continuous harassment by law enforcement officers. Among their many grievances, they complained about receiving so many traffic

citations that some were in danger of losing their driving privileges. During one lengthy discussion, we realized that all of them drove automobiles with Panther Party signs glued to their bumpers. This is a report of a study that I undertook to assess the seriousness of their charges and to determine whether we were hearing the voice of paranoia or reality. (Heussenstam, 1971, p. 32)

Our model does not start with this text but with its semantic representation, presented in Table 1. We follow here the conventions of Kintsch (1974) and Turner and Greene (in press). Propositions are numbered and are listed consecutively according to the order in which their predicates appear in the English text. Each proposition is enclosed by parentheses and contains a predicate (written first) plus one or more arguments. Arguments are concepts (printed in capital letters to distinguish them from words) or other embedded propositions, which are referred to by their number. Thus, as is shown in Table 1, (COMPLAIN, STUDENT, 19) is shorthand for (COMPLAIN, STUDENT, (HARASS, POLICE, STUDENT)). The semantic cases of the arguments have not been indicated in Table 1 to keep the notation simple, but they are obvious in most instances.

Some comments about this representation are in order. It is obviously very "surface" and, in connection with that, not at all unambiguous. This is a long way from a precise, logical formalism that would unambiguously represent "the meaning" of the text, preferably through some combination of elementary semantic predicates. The problem is not only that neither we nor anyone else has ever devised an adequate formalism of this kind; such a formalism would also be quite inappropriate for our purposes. If one wants to develop a psychological processing model of comprehension, one should start with a non-elaborated semantic representation that is close to the surface structure. It is then the task of the processing model to elaborate this representation in stages, just as a reader starts out with a conceptual analysis directly based on the surface structure but then constructs from this base more elaborate interpretations. To model some of these interpretative processes is the goal of the present enterprise. These processes change the initial semantic representation (e.g., by specifying coherence relations of a certain kind, adding required in-

Table 1
Proposition List for the Bumperstickers Paragraph

Proposition number	Proposition
1	(SERIES, ENCOUNTER)
2	(VIOLENT, ENCOUNTER)
3	(BLOODY, ENCOUNTER)
4	(BETWEEN, ENCOUNTER, POLICE, BLACK PANTHER)
5	(TIME: IN, ENCOUNTER, SUMMER)
6	(EARLY, SUMMER)
7	(TIME: IN, SUMMER, 1969)
8	(SOON, 9)
9	(AFTER, 4, 16)
10	(GROUP, STUDENT)
11	(BLACK, STUDENT)
12	(TEACH, SPEAKER, STUDENT)
13	(LOCATION: AT, 12, CAL STATE COLLEGE)
14	(LOCATION: AT, CAL STATE COLLEGE, LOS ANGELES)
15	(IS A, STUDENT, BLACK PANTHER)
16	(BEGIN, 17)
17	(COMPLAIN, STUDENT, 19)
18	(CONTINUOUS, 19)
19	(HARASS, POLICE, STUDENT)
20	(AMONG, COMPLAINT)
21	(MANY, COMPLAINT)
22	(COMPLAIN, STUDENT, 23)
23	(RECEIVE, STUDENT, TICKET)
24	(MANY, TICKET)
25	(CAUSE, 23, 27)
26	(SOME, STUDENT)
27	(IN DANGER OF, 26, 28)
28	(LOSE, 26, LICENSE)
29	(DURING, DISCUSSION, 32)
30	(LENGTHY, DISCUSSION)
31	(AND, STUDENT, SPEAKER)
32	(REALIZE, 31, 34)
33	(ALL, STUDENT)
34	(DRIVE, 33, AUTO)
35	(HAVE, AUTO, SIGN)
36	(BLACK PANTHER, SIGN)
37	(GLUED, SIGN, BUMPER)
38	(REPORT, SPEAKER, STUDY)
39	(DO, SPEAKER, STUDY)
40	(PURPOSE, STUDY, 41)
41	(ASSESS, STUDY, 42, 43)
42	(TRUE, 17)
43	(HEAR, 31, 44)
44	(OR, 45, 46)
45	(OF REALITY, VOICE)
46	(OF PARANOIA, VOICE)

Note. Lines indicate sentence boundaries. Propositions are numbered for ease of reference. Numbers as propositional arguments refer to the proposition with that number.

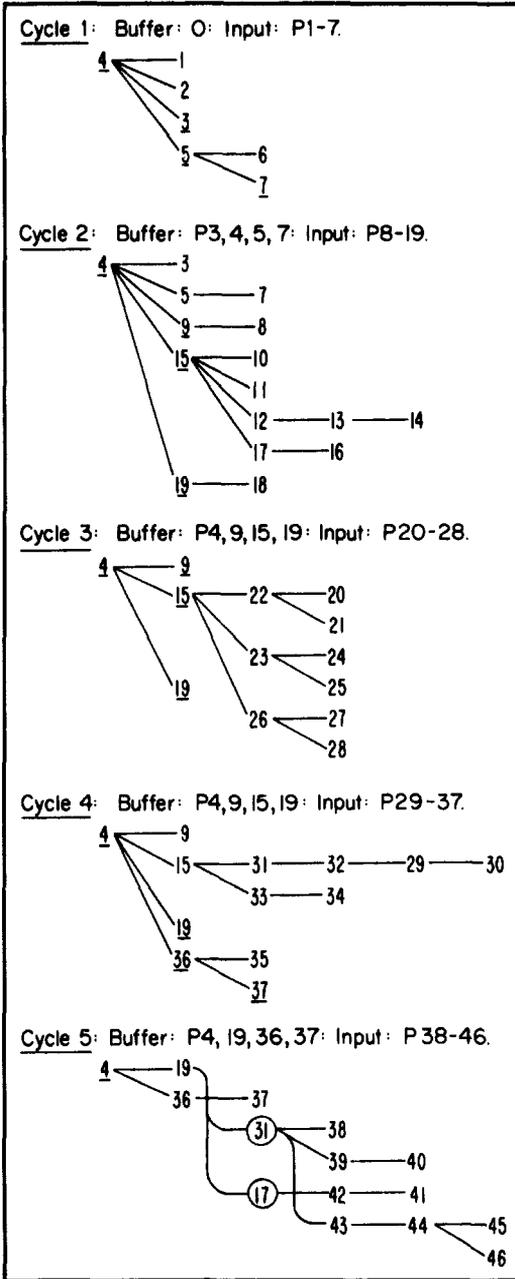


Figure 1. The cyclical construction of the coherence graph for the propositions (P) shown in Table 1.

ferences, and specifying a macrostructure) toward a deeper, less surface-dependent structure. Thus, the deep, elaborated semantic representation should stand at the end rather than at the beginning of a processing model, though even at that point some vagueness and

ambiguity may remain. In general, neither the original text nor a person's understanding of it is completely unambiguous.

This raises, of course, the question of what such a semantic representation represents. What is gained by writing (COMPLAIN, STUDENT) rather than *The students complained*? Capitalizing *complain* and calling it a concept rather than a word explains nothing in itself. There are two sets of reasons, however, why a model of comprehension must be based on a semantic representation rather than directly on English text. The notation clarifies which aspects of the text are important (semantic content) and which are not (e.g., surface structure) for our purpose; it provides a unit that appears appropriate for studying comprehension processes, that is, the proposition (see Kintsch & Keenan, 1973, for a demonstration that number of propositions in a sentence rather than number of words determines reading time); finally, this kind of notation greatly simplifies the scoring of experimental protocols in that it establishes fairly unambiguous equivalence classes within which paraphrases can be treated interchangeably.

However, there are other even more important reasons for the use of semantic representations. Work that goes beyond the confines of the present article, for example, on the organization of text bases in terms of fact frames and the generation of inferences from such knowledge fragments, requires the development of suitable representations. For instance, COMPLAIN is merely the name of a knowledge complex that specifies the normal uses of this concept, including antecedents and consequences. Inferences can be derived from this knowledge complex, for example, that the students were probably unhappy or angry or that they complained to someone, with the text supplying *the professor* for that role. Once a concept like COMPLAIN is elaborated in that way, the semantic notation is anything but vacuous. Although at present we are not concerned with this problem, the possibility for this kind of extension must be kept open.

Cycling

The first step in the processing model is to organize the input propositions into a coherent graph, as illustrated in Figure 1. It was

assumed for the purpose of this figure that the text is processed sentence by sentence. Since in our example the sentences are neither very short nor very long, this is a reasonable assumption and nothing more complicated is needed. It means that n_i , the number of input propositions processed per cycle, ranges between 7 and 12.

In Cycle 1 (see Figure 1), the buffer is empty, and the propositions derived from the first sentence are the input. P4 is selected as the superordinate proposition because it is the only proposition in the input set that is directly related to the title: It shares with the title the concept POLICE. (Without a title, propositions are organized in such a way that the simplest graph results.) P1, P2, P3, and P5 are directly subordinated to P4 because of the shared argument ENCOUNTER; P6 and P7 are subordinated to P5 because of the repetition of SUMMER.

At this point, we must specify the short-term memory assumptions of the model. For the present illustration, the short-term memory capacity was set at $s=4$. Although this seems like a reasonable value to try, there is no particular justification for it. Empirical methods to determine its adequacy will be described later. We also must specify some strategy for the selection of the propositions that are to be maintained in the short-term memory buffer from one cycle to the next. As was mentioned before, a good strategy should be biased in favor of superordinate and recent propositions. The "leading-edge strategy," originally proposed by Kintsch and Vipond (1978), does exactly that. It consists of the following scheme. Start with the top proposition in Figure 1 and pick up all propositions along the graph's lower edge, as long as each is more recent than the previous one (i.e., the index numbers increase); next, go to the highest level possible and pick propositions in order of their recency (i.e., highest numbers first); stop whenever s propositions have been selected. In Cycle 1, this means that the process first selects P4, then P5 and P7, which are along the leading edge of the graph; thereafter, it returns to Level 2 (Level 1 does not contain any other not-yet-selected propositions) and picks as a fourth and final proposition the most recent one from that level, that is, P3.

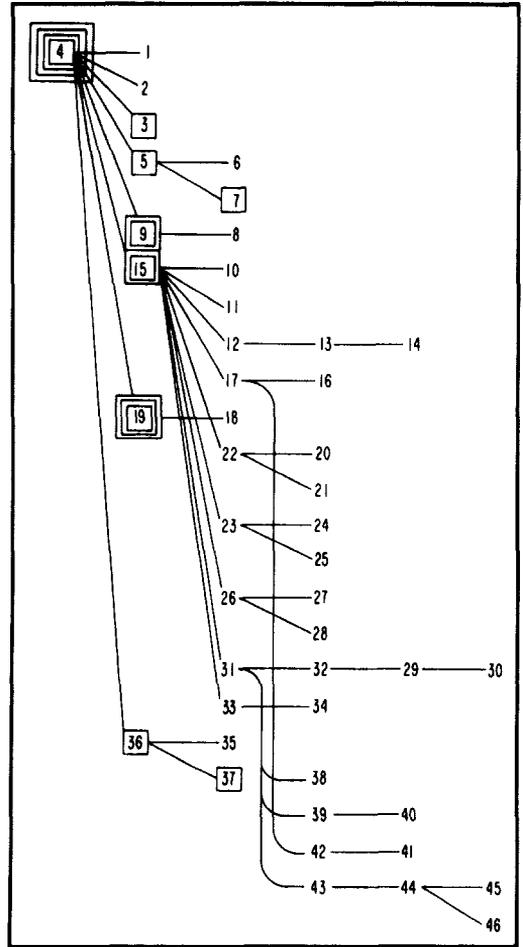


Figure 2. The complete coherence graph. (Numbers represent propositions. The number of boxes represents the number of extra cycles required for processing.)

Cycle 2 starts with the propositions carried over from Cycle 1; P9 is connected to P4, and P8 is connected to P9; the next connection is formed between P4 and P15 because of the common argument BLACK PANTHER; P19 also connects to P4 because of POLICE; P10, P11, P12, and P17 contain the argument STUDENT and are therefore connected to P15, which first introduced that argument. The construction of the rest of the graph as well as the other processing cycles should now be easy to follow. The only problems arise in Cycle 5, where the input propositions do not share a common argument with any of the propositions in the buffer. This requires a long-term memory search in this case; since some of the input

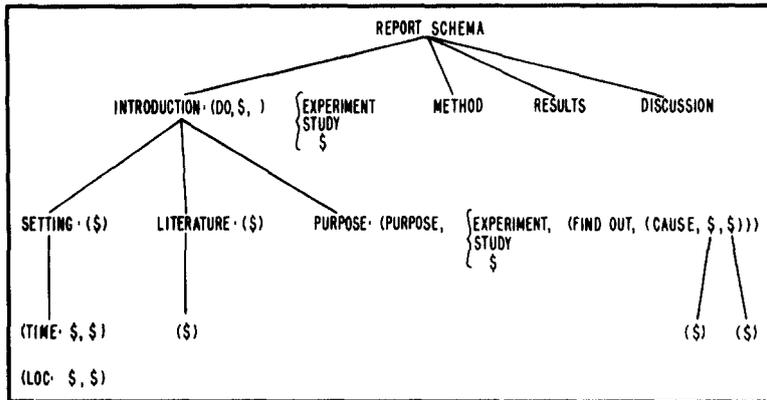


Figure 3. Some components of the report schema. (Wavy brackets indicate alternatives. \$ indicates unspecified information.)

propositions refer back to P17 and P31, the search leads to the reinstatement of these two propositions, as shown in Figure 1. Thus, the model predicts that (for the parameter values used) some processing difficulties should occur during the comprehension of the last sentence, having to do with determining the referents for "we" and "their charges."

The coherence graph arrived at by means of the processes illustrated in Figure 1 is shown in Figure 2. The important aspects of Figure 2 are that the graph is indeed connected, that is, the text is coherent, and that some propositions participated in more than one cycle. This extra processing is indicated in Figure 2 by the number of boxes in which each proposition is enclosed: P4 is enclosed by four boxes, meaning that it was maintained during four processing cycles after its own input cycle. Other propositions are enclosed by three, two, one, or zero boxes in a similar manner. This information is crucial for the model because it determines the likelihood that each proposition will be reproduced.

Schema

Figure 3 shows a fragment of the report schema. Since only the first paragraph of a report will be analyzed, only information in the schema relevant to the introduction of a report is shown. Reports are conventionally organized into introduction, method, results, and discussion sections. The introduction section must specify that an experiment, or

an observational study, or some other type of research was performed. The introduction contains three kinds of information: (a) the setting of the report, (b) literature references, and (c) a statement of the purpose (hypothesis) of the report. The setting category may further be broken down into location and time (\$ signs indicate unspecified information). The literature category is neglected here because it is not instantiated in the present example. The purpose category states that the purpose of the experiment is to find out whether some causal relation holds. The nature of this causal relation can then be further elaborated.

Macro-operations

In forming the macrostructure, micropropositions are either deleted, generalized, replaced by a construction, or carried over unchanged. Propositions that are not generalized are deleted if they are irrelevant; if they are relevant, they may become macropropositions. Whether propositions that are generalized become macropropositions depends on the relevance of the generalizations: Those that are relevant are included in the macrostructure with a higher probability than those that are not relevant. Thus, the first macro-operation is to form all generalizations of the micropropositions. Since the present theory lacks a formal inference component, intuition must once again be invoked to provide us with the required generalizations. Table 2 lists the generalizations that we have noted in the text

base of the Bumperstickers paragraph. They are printed below their respective micropropositions and indicated by an "M." Thus, P1 (SERIES, ENCOUNTER) is generalized to M1 (SOME, ENCOUNTER).

In Figure 4, the report schema is applied to our text-base example in order to determine which of the propositions are relevant and which are not. The schema picks out the generalized setting statements—that the whole episode took place *in California, at a college, and in the sixties*. Furthermore, it selects that *an experiment was done with the purpose of finding out whether Black Panther bumperstickers were the cause of students receiving traffic tickets*. Associated with the bumperstickers is the fact that the *students had cars with the signs on them*. Associated with the tickets is *many tickets*, and that *the students complained about police harassment*. Once these propositions have been determined as relevant, a stricter relevance criterion selects not all generalized setting information, but only the major setting (*in California*), and not all antecedents and consequences of the basic causal relationship, but only its major components (*bumperstickers lead to tickets*). Finally, at the top level of the macrostructure hierarchy, all the information in the introduction is reduced to *an experiment was done*. Thus, some propositions appear at only one level of the hierarchy, others at two levels, and one at all three levels. Each time a proposition is selected at a particular level of the macrostructure, the likelihood that it will later be recalled increases.

The results of both the micro- and macroprocesses are shown in Table 2. This table contains the 46 micropropositions shown in Table 1 and their generalizations. Macro-

propositions are denoted by M in the table; micropropositions that are determined to be relevant by the schema and thus also function as macropropositions are denoted by MP.

The reproduction probabilities for all these propositions are also derived in Table 2. Three kinds of storage operations are distinguished. The operation S is applied every time a (micro-) proposition participates in one of the cycles of Figure 1, as shown by the number of boxes in Figure 2. Different operators apply to macropropositions, depending on whether or not they are relevant. If a macroproposition is relevant, an operator M applies; if it is irrelevant, the operator is G. Thus, the reproduction probability for the irrelevant macroproposition *some encounters* is determined by G; but for the relevant *in the sixties*, it is determined by M. Some propositions can be stored either as micro- or macropropositions; for example, the operator SM² is applied to MP23 — S because P23 participates in one processing cycle and M² because M23 is incorporated in two levels of the macrostructure.

The reproduction probabilities shown in the last column of Table 2 are directly determined by the third column of the table, with *p*, *g*, and *m* being the probabilities that each application of S, G, and M, respectively, results in a successful reproduction of that proposition. It is assumed here that S and M are statistically independent.

Production

Output protocols generated by the model are illustrated in Tables 3 and 4. Strictly speaking, only the reproduced text bases shown in these tables are generated by the

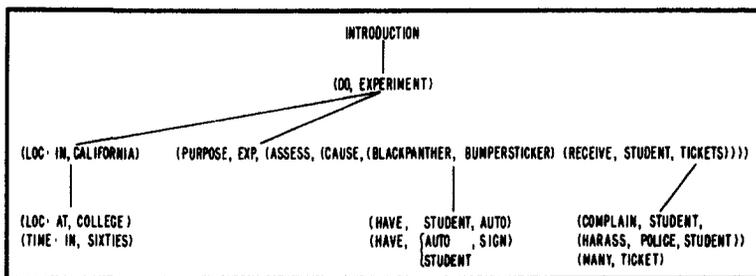


Figure 4. The macrostructure of the text base shown in Table 1. (Wavy brackets indicate alternatives.)

model; this output was simulated by reproducing the propositions of Table 2 with the probabilities indicated. (The particular values of p , g , and m used here appear reasonable in light of the data analyses to be reported in the next section.) The results shown are from a single simulation run, so that no selection is involved. For easier reading, English sentences instead of semantic representations are shown.

The simulated protocols contain some meta-statements and reconstructions, in addition to the reproduced propositional content. Nothing in the model permits one to assign a probability value to either metastatements or reconstructions. All we can say is that such statements occur, and we can identify them as metastatements or reconstructions if they are encountered in a protocol.

Tables 3 and 4 show that the model can produce recall and summarization protocols that, with the addition of some metastatements and reproductions, are probably indistinguishable from actual protocols obtained in psychological experiments.

Preliminary Data Analyses

The goal of the present section is to demonstrate how actual experimental protocols can be analyzed with the methods developed here. We have used "Bumperstickers and the Cops" in various experiments. The basic procedure of all these experiments was the same, and only the length of the retention interval varied. Subjects read the typewritten text at their own speed. Thereafter, subjects were asked to recall the whole report, as well as they could, not necessarily verbatim. They were urged to keep trying, to go over their protocols several times, and to add anything that came to mind later. Most subjects worked for at least 20 minutes at this task, some well over an hour. The subjects typed their protocols into a computer-controlled screen and were shown how to use the computer to edit and change their protocols. The computer recorded their writing times. After finishing the recall protocol, a subject was asked to write a summary of the report. The summary had to be

Table 2
Memory Storage of Micro- and Macropropositions

Proposition number	Proposition	Storage operation	Reproduction probability
P1	(SERIES, ENC)	S	p
M1	(SOME, ENC)	G	g
P2	(VIOL, ENC)	S	p
P3	(BLOODY, ENC)	S^2	$1 - (1 - p)^2$
P4	(BETW, ENC, POL, BP)	S^5	$1 - (1 - p)^5$
P5	(IN, ENC, SUM)	S^2	$1 - (1 - p)^2$
P6	(EARLY, SUM)	S	p
P7	(IN, SUM, 1969)	S^2	$1 - (1 - p)^2$
M7	(IN, EPISODE, SIXTIES)	M	m
P8	(SOON, 9)	S	p
P9	(AFTER, 4, 16)	S^3	$1 - (1 - p)^3$
P10	(GROUP, STUD)	S	p
M10	(SOME, STUD)	G	g
P11	(BLACK, STUD)	S	p
P12	(TEACH, SPEAK, STUD)	S	p
M12	(HAVE, SPEAK, STUD)	G	g
P13	(AT, 12, CSC)	S	p
M13	(AT, EPISODE, COLLEGE)	M	m
P14	(IN, CSC, LA)	S	p
M14	(IN, EPISODE, CALIF)	M^2	$1 - (1 - m)^2$
P15	(IS A, STUD, BP)	S^3	$1 - (1 - p)^3$
P16	(BEGIN, 17)	S	p
MP17	(COMPLAIN, STUD, 19)	SM	$p + m - pm$
P18	(CONT, 19)	S	p
MP19	(HARASS, POL, STUD)	S^4M	$1 - (1 - p)^4 + m - [1 - (1 - p)^4]m$

Table 2 (continued)

Proposition number	Proposition	Storage operation	Reproduction probability
P20	(AMONG, COMP)	S	p
P21	(MANY, COMP)	S	p
P22	(COMPLAIN, STUD, 23)	S	p
MP23	(RECEIVE, STUD, TICK)	SM ²	$p + 1 - (1 - m)^2 - p[1 - (1 - m)^2]$
MP24	(MANY, TICK)	SM	$p + m - pm$
P25	(CAUSE, 23, 27)	S	p
P26	(SOME, STUD)	S	p
P27	(IN DANGER, 26, 28)	S	p
P28	(LOSE, 26, LIC)	S	p
<hr/>			
P29	(DURING, DISC)	S	p
P30	(LENGTHY, DISC)	S	p
P31	(AND, STUD, SPEAK)	S	p
P32	(REALIZE, 31, 34)	S	p
P33	(ALL, STUD)	S	p
P34	(DRIVE, 33, AUTO)	S	p
M34	(HAVE, STUD, AUTO)	M	m
MP35	(HAVE, AUTO/STUD, SIGN)	SM	$p + m - pm$
MP36	(BP, SIGN)	S ² M ²	$1 - (1 - p)^2 + 1 - (1 - m)^2 - [1 - (1 - p)^2] \times [1 - (1 - m)^2]$
P37	(GLUED, SIGN, BUMP)	S	p
M37	(ON, SIGN, BUMP)	M ²	$1 - (1 - m)^2$
<hr/>			
P38	(REPORT, SPEAK, STUDY)	S	p
MP39	(DO, SPEAK, STUDY)	SM ³	$p + 1 - (1 - m)^3 - p[1 - (1 - m)^3]$
MP40	(PURPOSE, STUDY, 41)	SM ²	$p + 1 - (1 - m)^2 - p[1 - (1 - m)^2]$
MP41	(ASSESS, STUDY, 42, 43)	SM ²	$p + 1 - (1 - m)^2 - p[1 - (1 - m)^2]$
MP42	(TRUE, 17)/(CAUSE, 36, 23)	SM ²	$p + 1 - (1 - m)^2 - p[1 - (1 - m)^2]$
P43	(HEAR, 31, 44)	S	p
P44	(OR, 45, 46)	S	p
P45	(OF REALITY, VOICE)	S	p
P46	(OF PARANOIA, VOICE)	S	p

Note. Lines show sentence boundaries. P indicates a microproposition, M indicates a macroproposition, and MP is both a micro- and macroproposition. S is the storage operator that is applied to micropropositions, G is applied to irrelevant macropropositions, and M is applied to relevant macropropositions.

between 60 and 80 words long to facilitate between-subjects comparisons. The computer indicated the number of words written to avoid the distraction of repeated word counts. All subjects were undergraduates from the University of Colorado and were fulfilling a course requirement.

A group of 31 subjects was tested immediately after reading the report. Another group of 32 subjects was tested after a 1-month retention interval, and 24 further subjects were tested after 3 months.

Some statistics of the protocols thus obtained are shown in Table 5. Since the summaries were restricted in length, the fact that their average length did not change appreciably

over a 3-month period is hardly remarkable. But note that the overall length of the free-recall protocols declined only by about a third or quarter with delay. Obviously, subjects did very well even after 3 months, at least in terms of the number of words produced. If we restrict ourselves to the first paragraph of the text and analyze propositionally those portions of the recall protocols referring to that paragraph, we arrive at the same conclusions. Although there is a (statistically significant) decline over delay in the number of propositions in the protocols that could be assigned to the introduction category, the decline is rather moderate.

While the output changed relatively little

Table 3
A Simulated Recall Protocol

Reproduced text base
P1, P4, M7, P9, P13, P17, MP19, MP35, MP36, M37, MP39, MP40, MP42
Text base converted to English, with reconstructions italicized
In the sixties (M7), there was a series (P1) of riots between the police and the Black Panthers (P4). <i>The police did not like the Black Panthers</i> (normal consequence of 4). After that (P9), students <i>who were enrolled</i> (normal component of 13) at the California State College (P13) complained (P17) that the police were harassing them (P19). <i>They were stopped and had to show their driver's license</i> (normal component of 23). The police gave them trouble (P19) because (MP42) they had (MP35) Black Panther (MP36) signs on their bumpers (M37). <i>The police discriminated against these students and violated their civil rights</i> (particularization of 42). They did an experiment (MP39) for this purpose (MP40).

Note. The parameter values used were $p = .10$, $g = .20$, and $m = .40$. P, M, and MP identify propositions from Table 2.

in quantity, its qualitative composition changed a great deal. In the recall protocols, the proportion of reproductive propositions in the total output declined from 72% to 48%, while at the same time, reconstructions almost doubled their contribution, and metastatements quadrupled theirs.⁴ These results are shown in Figure 5; they are highly significant statistically, $\chi^2(4) = 56.43$. As less material from the text is reproduced, proportionately more material is added by the production processes themselves. Interestingly, these production processes operated quite accurately in the present case, even after 3 months: Errors, that is, wrong guesses or far-fetched confabulations, occurred with negligible frequency at all delay intervals (1%, 0%, and 1% for zero, 1-month, and 3-month delays, respectively).

Table 4
A Simulated Summary

Reproduced text base
M14, M37, MP39, MP40
Text base converted to English, with metastatements and reconstructions italicized
<i>This report was about</i> (metastatement) an experiment done (MP39) in California (M14). It was done (MP39) in order to (MP40) <i>protect students from discrimination</i> (normal consequence of 41).

Note. The parameter values used were $p = .01$, $g = .05$, and $m = .30$. M and MP identify propositions from Table 2.

For the summaries, the picture is similar, though the changes are not quite as pronounced, as shown in Figure 6. The contribution of the three response categories to the total output still differs significantly, however, $\chi^2(4) = 16.90$, $p = .002$. Not surprisingly, summaries are generally less reconstructive than recall. Note that the reproductive component includes reproductions of both micro- and macropropositions!

We have scored every statement that could be a reproduction of either a macro- or micro-proposition as if it actually were. This is, of course, not necessary: It is not only possible but quite probable that some of the statements that we regarded as reproductive in origin are reconstructions that accidentally duplicate actual text propositions. Thus, our estimates of the reproductive components of recall and summarization protocols are in fact upper bounds.

Protocol Analyses

In order to obtain the data reported above, all protocols had to be analyzed proposi-

⁴Metastatements seem to be most probable when one has little else to say: Children use such statements most frequently when they talk about something they do not understand (e.g., Poulsen et al., in press), and the speech of aphasics is typically characterized by such semantically empty comments (e.g., Luria & Tsvetkova, 1968). In both cases, the function of these statements may be to maintain the communicative relationship between speaker and listener. Thus, their function is a pragmatic one.

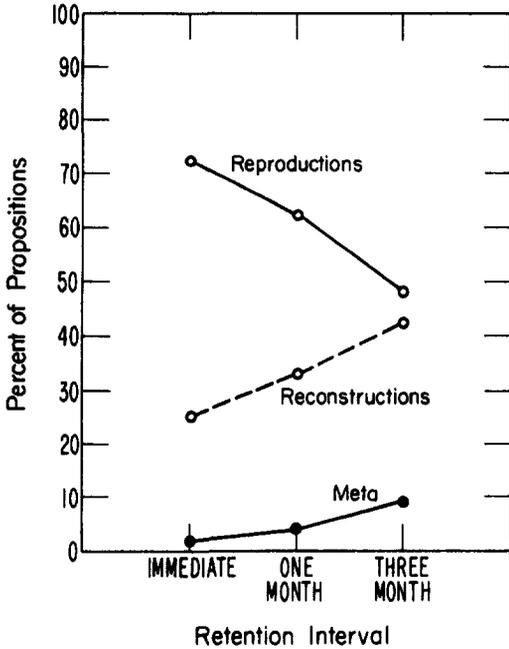


Figure 5. Proportion of reproductions, reconstructions, and metastatements in the recall protocols for three retention intervals.

tionally. An example of how this kind of analysis works in detail is presented in Tables 6 and 7, for which one subject from the immediate condition was picked at random. The recall protocol and the summary of this subject are shown, together with the propositional representations of these texts. In constructing this representation, an effort was made to bring out any overlap between the propositional representation of the original text and its macrostructure (see Table 2) and the protocol. In other words, if the protocol contained a statement that was semantically (not necessarily lexically) equivalent to a text proposition, this text proposition was used in the analysis. This was done to simplify the analysis; as mentioned before, a fine-grained analysis distinguishing optional meaning-preserving transformations would be possible.

Each protocol proposition was then assigned to one of four response categories: (a) reproductions (the index number of the reproduced proposition is indicated), (b) reconstructions (the source of the reconstruction is noted wherever possible), (c) metastatements, and

Table 5
Average Number of Words and Propositions of Recall and Summarization Protocols as a Function of the Retention Interval

Protocol and retention interval	No. words (total protocol)	No. propositions (1st paragraph only)
Recall		
Immediate	363	17.3
1 month	225	13.2
3 months	268	14.8
Summaries		
Immediate	75	8.4
1 month	76	7.4
3 months	72	6.2

(d) errors and unclassifiable statements (not represented in the present protocol).

Note that the summary is, for the most part, a reduced version of the recall protocol; on the other hand, the second proposition of the summary that specifies the major setting did not appear in the recall. Clearly, the recall

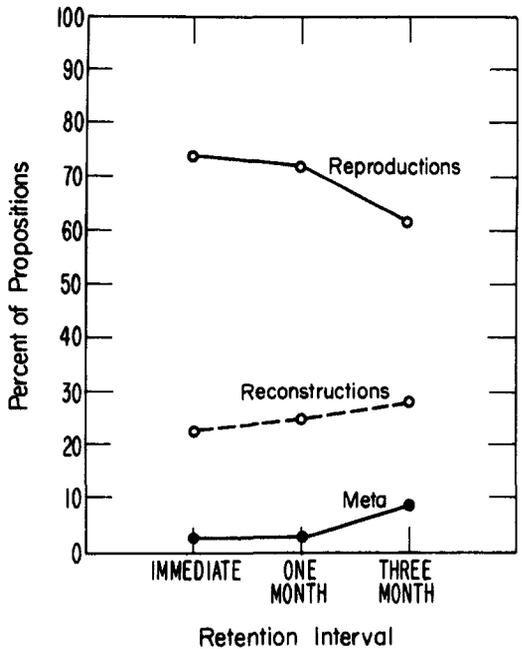


Figure 6. Proportion of reproductions, reconstructions, and metastatements in the summaries for three retention intervals.

Table 6

Recall Protocol and Its Analysis From a Randomly Selected Subject

Recall protocol		
Proposition number	Proposition	Classification
1	(START WITH, REPORT, 2)	Schematic metastatement
2	(TELL, REPORT, BLACK PANTHER)	Semantic metastatement
3	(TIME: IN, BLACK PANTHER, SIXTIES)	M7
4	(TELL, PROFESSOR, STUDENT)	Semantic metastatement
5	(SOME, STUDENT)	M10
6	(HAVE, PROFESSOR, STUDENT)	M12
7	(IS A, STUDENT, BLACK PANTHER)	P15 (lexical transformation)
8	(COMPLAIN, STUDENT, 9)	MP17
9	(RECEIVE, STUDENT, TICKET, POLICE)	MP23 (specification of component)
10	(FEEL, STUDENT)	Specification of condition of MP17
11	(CAUSE, 7, 9)	MP42
12	(DECIDE, PROFESSOR, 13)	Specification of condition of MP39
13	(DO, PROFESSOR, EXPERIMENT)	MP39
14	(PURPOSE, 13, 15)	MP40
15	(ASSESS, 16)	MP41
16	(CAUSE, 17, 9)	MP42
17	(ON, BUMPER, SIGN)	M37
18	(HAVE, AUTO, SIGN)	MP35

Note. Only the portion referring to the first paragraph of the text is shown. Lines indicate sentence boundaries. P, M, and MP identify propositions from Table 2.

protocol does not reflect only what is stored in memory.

Statistical Analyses

While metastatements and reconstructions cannot be analyzed further, the reproductions permit a more powerful test of the model. If one looks at the pattern of frequencies with which both micro- and macropropositions (i.e., Table 2) are reproduced in the various recall conditions, one notes that although the three delay conditions differ among each other significantly [minimum $t(54) = 3.67$], they are highly correlated: Immediate recall correlates .88 with 1-month recall and .77 with 3-month recall; the two delayed recalls correlate .97 between each other. The summaries are even more alike at

the three retention intervals. They do not differ significantly [maximum $t(54) = 1.99$], and the intercorrelations are .92 for immediate/1-month, .87 for immediate/3-month, and .94 for 1-month/3-month recall. These high correlations are not very instructive, however, because they hide some subtle but consistent changes in the reproduction probabilities. A more powerful analysis that brings out these changes is provided by the model.

The predicted frequencies for each micro- and macroproposition are shown in Table 3 as a function of the parameters p , g , and m . These equations are, of course, for a special case of the model, assuming chunking by sentences, a buffer capacity of four propositions, and the leading-edge strategy. It is, nevertheless, worthwhile to see how well this special case of the model can fit the data ob-

Table 7

Summary and Its Analysis From the Same Randomly Selected Subject as in Table 6

Summary		
Proposition number	Proposition	Classification
1	(IS ABOUT, REPORT, EXPERIMENT)	Semantic metastatement
2	(LOC: IN, 1, LOS ANGELES)	M14
3	(PURPOSE, EXPERIMENT, 4)	MP40
4	(CAUSE, 5, 6)	MP42
5	(IS A, SOMEONE, BLACK PANTHER)	P15
6	(GIVE, POLICE, TICKET, SOMEONE)	MP23 (lexical transformation)

Note. Only the portion referring to the first paragraph of the text is shown. P, M, and MP identify propositions from Table 2.

tained from the first paragraph of Bumperstickers when the three statistical parameters are estimated from the data. This was done by means of the STEPIT program (Chandler, 1965), which for a given set of experimental frequencies, searches for those parameter values that minimize the chi-squares between predicted and obtained values. Estimates obtained for six separate data sets (three retention intervals for both recall and summaries) are shown in Table 8. We first note that the goodness of fit of the special case of the model under consideration here was quite good in five of the six cases, with the obtained values being slightly less than the expected values.⁵ Only the immediate recall data

could not be adequately fit by the model in its present form, though even there the numerical value of the obtained goodness-of-fit statistic is by no means excessive. (The problems are caused mostly by two data points: Subjects produced P11 and P33 with greater frequency than the model could account for; this did not happen in any of the other data sets.)

Perhaps more important than the goodness-of-fit statistics is the fact that the pattern of parameter estimates obtained is a sensible and instructive one: In immediate recall, the probability of reproducing micropropositions (p) is about five times as high as after 3 months; the probability of reproducing irrelevant generalizations (g) decreases by a similar factor; however, the likelihood that macropropositions are reproduced (m) changes very little with delay. In spite of the high correlations observed between the immediate recall and 3-month delayed recall, the forgetting rates for micropropositions appear to be about four times greater than that for macropropositions! In the summaries, the same kind of changes occur, but they are rather less pronounced. Indeed, the summary data from all three delay intervals could be pooled and fitted

Table 8
Parameter Estimates and Chi-Square
Goodness-of-Fit Values

Protocol and retention interval	\hat{p}	\hat{m}	\hat{g}	$\chi^2(50)$
Recall				
Immediate	.099	.391	.198	88.34*
1 month	.032	.309	.097	46.08
3 months	.018	.321	.041	37.41
Summaries				
Immediate	.015	.276	.026	30.11
1 month	.024	.196	.040	46.46
3 months	.001	.164	.024	30.57

Note. The average standard errors are .003 for \hat{p} , .023 for \hat{m} , and .025 for \hat{g} , respectively.

* $p < .001$.

⁵ Increasing (or decreasing) all parameter values in Table 8 by 33% yields statistically significant increases in the chi-squares. On the other hand, the increases in the chi-squares that result from 10% changes in the parameter values are not sufficiently large to be statistically significant.

by one set of parameter values with little effect on the goodness of fit. Furthermore, recall after 3 months looks remarkably like the summaries in terms of the parameter estimates obtained here.

These observations can be sharpened considerably through the use of a statistical technique originally developed by Neyman (1949). Neyman showed that if one fits a model with some set of r parameters, yielding a $\chi^2(n-r)$, and then fits the same data set again but with a restricted parameter set of size q , where $q < r$, and obtains a $\chi^2(n-q)$, $\chi^2(n-r) - \chi^2(n-q)$ also has the chi-square distribution with $r-q$ degrees of freedom under the hypothesis that the $r-q$ extra parameters are merely capitalizing on chance variations in the data, and that the restricted model fits the data as well as the model with more parameters. Thus, in our case, one can ask, for instance, whether all six sets of recall and summarization data can be fitted with one set of the parameters p , m , and g , or whether separate parameter values for each of the six conditions are required. The answer is clearly that the data cannot be fitted with one single set of parameters, with the chi-square for the improvement obtained by separate fits, $\chi^2(15) = 272.38$. Similarly, there is no single set of parameter values that will fit the three recall conditions jointly, $\chi^2(6) = 100.2$. However, single values of p , m , and g will do almost as well as three separate sets for the summary data. The improvement from separate sets is still significant, $\chi^2(6) = 22.79$, $.01 < p < .001$, but much smaller in absolute terms (this corresponds to a normal deviate of 3.1, while the previous two chi-squares translate into z scores of 16.9 and 10.6, respectively). Indeed, it is almost possible to fit jointly all three sets of summaries and the 3-month delay recall data, though the resulting chi-square for the significance of the improvement is again significant statistically, $\chi^2(9) = 35.77$, $p < .001$. The estimates that were obtained when all three sets of summaries were fitted with a single set of parameters were $\hat{p} = .018$, $\hat{m} = .219$, and $\hat{g} = .032$; when the 3-month recall data were included, these values changed to $\hat{p} = .018$, $\hat{m} = .243$, and $\hat{g} = .034$.

The summary data can thus be described in terms of the model by saying that over 70%

of the protocols consisted of reproductions, with the remainder being mostly reconstructions and a few metastatements. With delay, the reproductive portion of these protocols decreased but only slightly. Most of the reproductions consisted of macropropositions, with the reproduction probability for macropropositions being about 12 times that for micropropositions. For recall, there were much more dramatic changes with delay: For immediate recall, reproductions were three times as frequent as reconstructions, while their contributions are nearly equal after 3 months. Furthermore, the composition of the reproduced information changes very much: Macropropositions are about four times as important as micropropositions in immediate recall; however, that ratio increases with delay, so after 3 months, recall protocols are very much like summaries in that respect.

It is interesting to contrast these results with data obtained in a different experiment. Thirty subjects read only the first paragraph of Bumperstickers and recalled it in writing immediately. The macroprocesses described by the present model are clearly inappropriate for these subjects: There is no way these subjects can construct a macrostructure as outlined above, since they do not even know that what they are reading is part of a research report! Hence, one would expect the pattern of recall to be distinctly different from that obtained above; the model should not fit the data as well as before, and the best-fitting parameter estimates should not discriminate between macro- and micropropositions. Furthermore, in line with earlier observations (e.g., Kintsch et al., 1975), one would expect most of the responses to be reproductive. This is precisely what happened: 87% of the protocols consisted of reproductions, 10% reconstructions, 2% metastatements, and 1% errors. The total number of propositions produced was only slightly greater (19.9) than immediate recall of the whole text (17.3; see Table 5). The correlation with immediate recall of the whole text was down to $r = .55$; with the summary data, there was even less correlation, for example, $r = .32$, for the 3-month delayed summary. The model fit was poor, with a minimum chi-square of 163.72 for 50 df . Most interesting, however, is the

fact that the estimates for the model parameters p , m , and g that were obtained were not differentiated: $\hat{p} = .33$, $\hat{m} = .30$, $\hat{g} = .36$. Thus, micro- and macroinformation is treated alike here—very much unlike the situation where subjects read a long text and engage in schema-controlled macro-operations!

While the tests of the model reported here certainly are encouraging, they fall far short of a serious empirical evaluation. For this purpose, one would have to work with at least several different texts and different schemata. Furthermore, in order to obtain reasonably stable frequency estimates, many more protocols should be analyzed. Finally, the general model must be evaluated, not merely a special case that involves some quite arbitrary assumptions.⁶ However, once an appropriate data set is available, systematic tests of these assumptions become feasible. Should the model pass these tests satisfactorily, it could become a useful tool in the investigation of substantive questions about text comprehension and production.

General Discussion

A processing model has been presented here for several aspects of text comprehension and production. Specifically, we have been concerned with problems of text cohesion and gist formation as components of comprehension processes and the generation of recall and summarization protocols as output processes. According to this model, coherent text bases are constructed by a process operating in cycles and constrained by limitations of working memory. Macroprocesses are described that reduce the information in a text base through deletion and various types of inferences to its gist. These processes are under the control of the comprehender's goal, which is formalized in the present model as a schema. The macroprocesses described here are predictable only if the controlling schema can be made explicit. Production is conceived both as a process of reproducing stored text information, which includes the macrostructure of the text, and as a process of construction: Plausible inferences can be generated through the inverse application of the macro-operators.

This processing model was derived from the considerations about semantic structures

sketched at the beginning of this article. There, we limited our discussion to points that were directly relevant to our processing model. The model, as developed here, is, however, not a comprehensive one, and it now becomes necessary to extend the discussion of semantic structures beyond the self-imposed limits we have observed so far. This is necessary for two reasons: (a) to indicate the directions that further elaborations of the model might take and (b) to place what we have so far into a broader context.

Coherence

Sentences are assigned meaning and reference not only on the basis of the meaning and reference of their constituent components but also relative to the interpretation of other, mostly previous, sentences. Thus, each sentence or clause is subject to contextual interpretation. The cognitive correlate of this observation is that a language user needs to relate new incoming information to the information he or she already has, either from the text, the context, or from the language user's general knowledge system. We have modeled this process in terms of a coherent propositional network: The referential identity of concepts was taken as the basis of the co-

⁶ We have used the leading-edge strategy here because it was used previously by Kintsch and Vipond (1978). In comparison, for the data from the immediate-summary condition, a recency-only strategy increases the minimum chi-square by 43%, while a levels-plus-primacy strategy yields an increase of 23% (both are highly significant). A random selection strategy can be rejected unequivocally, $\chi^2(34) = 113.77$. Of course, other plausible strategies must be investigated, and it is quite possible that a better strategy than the leading-edge strategy can be found. However, a systematic investigation would be pointless with only a single text sample.

Similar considerations apply to the determination of the buffer capacity, s . The statistics reported above were calculated for an arbitrarily chosen value of $s = 4$. For $s = 10$, the minimum chi-square increases drastically, and the model no longer can fit the data. On the other hand, for $s = 1, 2, \text{ or } 3$, the minimum chi-squares are only slightly larger than those reported in Table 8 (though many more reinstatement searches are required during comprehension for the smaller values of s). A more precise determination of the buffer capacity must be postponed until a broader data base becomes available.

herence relationships among the propositions of a text base. While referential coherence is undoubtedly important psychologically, there are other considerations that need to be explicated in a comprehensive theory of text processing.

A number of experimental studies, following Haviland and Clark (1974), demonstrate very nicely the role that referential coherence among sentences plays in comprehension. However, the theoretical orientation of these studies is somewhat broader than the one we have adopted above. It derives from the work of the linguists of the Prague School (e.g., Sgall & Hajičová, 1977) and others (e.g., Halliday, 1967) who are concerned with the given-new articulation of sentences, as it is described in terms of topic-comment and presupposition-assertion relations. Within psychology, this viewpoint has been expressed most clearly by Clark (e.g., Clark, 1977). Clark's treatment of the role of the "given-new contract" in comprehension is quite consistent with the processing model offered here. Referential coherence has an important function in Clark's system, but his notions are not limited to it, pointing the way to possible extensions of the present model.

Of the two kinds of semantic relations that exist among propositions, extensional and intensional ones, we have so far restricted ourselves to the former. That is, we have related propositions on the basis of their reference rather than their meaning. In the predicate logical notation used here, this kind of referential relation is indicated by common arguments in the respective propositions. In natural surface structures, referential identity is most clearly expressed by pronouns, other "pro" forms, and definite articles. Note that referential identity is not limited to individuals, such as discrete objects, but may also be based on related sets of propositions specifying properties, relations, and events, or states of affairs. Identical referents need not be referred to with the same expressions: We may use the expressions "my brother," "he," "this man," or "this teacher" to refer to the same individual, depending on which property of the individual is relevant in the respective sentences of the discourse.

Although referential identity or other refer-

ential relationships are frequent concomitant properties of coherence relations between propositions of a text base, such relations are not sufficient and sometimes not even necessary. If we are talking about John, for instance, we may say that he was born in London, that he was ill yesterday, or that he now smokes a pipe. In that case, the referential individual remains identical, but we would hardly say that a discourse mentioning these facts about him would by this condition alone be coherent. It is essential that the facts themselves be related. Since John may well be a participant in such facts, referential identity may be a normal feature of coherence as it is established between the propositions taken as denoting facts.

Facts

The present processing model must therefore be extended by introducing the notion of "facts." Specifically, we must show how the propositions of a text base are organized into higher-order fact units. We shall discuss here some of the linguistic theory that underlies these notions.

The propositions of a text base are connected if the facts denoted by them are related. These relations between facts in some possible world (or in related possible worlds) are typically of a conditional nature, where the conditional relation may range from possibility, compatibility, or enablement via probability to various kinds of necessity. Thus, one fact may be a possible, likely, or necessary condition of another fact; or a fact may be a possible, likely, or necessary consequence of another fact. These connection relations between propositions in a coherent text base are typically expressed by connectives such as "and," "but," "because," "although," "yet," "then," "next," and so on.

Thus, facts are joined through temporal ordering or presuppositional relationships to form more complex units of information. At the same time, facts have their own internal structure that is reflected in the case structure of clauses and sentences. This framelike structure is not unlike that of propositions, except that it is more complex, with slots that assign specific roles to concepts, propositions, and other facts. The usual account of such

structures at the level of sentence meaning is given in terms of the following:

1. state/event/action (predicate)
2. participants involved (arguments)
 - a. agent(s)
 - b. patient(s)
 - c. beneficiary
 - d. object(s)
 - e. instrument(s)
and so on
3. circumstantials
 - a. time
 - b. place
 - c. direction
 - d. origin
 - e. goal
and so on
4. properties of 1, 2, and 3
5. modes, moods, modalities.

Without wanting to be more explicit or complete, we surmise that a fact unit would feature categories such as those mentioned above. The syntactic and lexical structure of the clause permits a first provisional assignment of such a structure. The interpretation of subsequent clauses and sentences might correct this hypothesis.

In other words, we provisionally assume that the case structure of sentences may have processing relevance in the construction of complex units of information. The propositions play the role of characterizing the various properties of such a structure, for example, specifying which agents are involved, what their respective properties are, how they are related, and so on. In fact, language users are able to derive the constituent propositions from once-established fact frames.

We are now in a position to better understand how the content of the proposition in a text base is related, namely, as various kinds of relations between the respective individuals, properties, and so on characterizing the connected facts. The presupposition-assertion structure of the sequence shows how new facts are added to the old facts; more specifically, each sentence may pick out one particular aspect, for example, an individual or property already introduced before, and assign it further properties or introduce new individuals related to individuals introduced before (see Clark, 1977; van Dijk, 1977d).

It is possible to characterize relations be-

tween propositions in a discourse in more functional terms, at the pragmatic or rhetorical levels of description. Thus, one proposition may be taken as a "specification," a "correction," an "explanation," or a "generalization" of another proposition (Meyer, 1975; van Dijk, 1977c, 1977d). However, we do not yet have an adequate theory of such functional relations.

The present model was not extended beyond the processes involved in referential coherence of texts because we do not feel that the problems involved are sufficiently well understood. However, by limiting the processing model to coherence in terms of argument repetition, we are neglecting the important role that fact relationships play in comprehension. For instance, most of the inferences that occur during comprehension probably derive from the organization of the text base into facts that are matched up with knowledge frames stored in long-term memory, thus providing information missing in the text base by a process of pattern completion, or "default assignments," in Minsky's terminology (Minsky, 1975). Likewise, the fate of inconsistent information might be of interest in a text, for example, whether it would be disregarded in recall or, on the contrary, be particularly well reported because of the extra processing it receives. Finally, by looking at the organization of a text base in terms of fact relations, a level of representation is obtained that corresponds to the "knowledge structures" of Schank and Abelson (1977). Schank and Abelson have argued convincingly that such a level of representation is necessary in order to account for full comprehension. Note, however, that in their system, as in the present one, that level of representation is built on earlier levels (their microscopic and macroscopic conceptual dependency representations, which correspond loosely to our microstructure and macrostructure).

Limitations and Future Directions

Other simplifying assumptions made in the course of developing this model should likewise be examined. Consider first the relationship between a linguistic text grammar and a text base actually constructed by a reader or listener. In general, the reader's text base will be different from the one abstractly specified

by a text grammar. The operations and strategies involved in discourse understanding are not identical to the theoretical rules and constraints used to generate coherent text bases. One of the reasons for these differences lies in the resource and memory limitations of cognitive processes. Furthermore, in a specific pragmatic and communicative context, a language user is able to establish textual coherence even in the absence of strict, formal coherence. He or she may accept discourses that are not properly coherent but that are nevertheless appropriate from a pragmatic or communicative point of view. Yet, in the present article, we have neglected those properties of natural language understanding and adopted the working hypothesis that full understanding of a discourse is possible at the semantic level alone. In future work, a discourse must be taken as a complex structure of speech acts in a pragmatic and social context, which requires what may be called *pragmatic comprehension* (see van Dijk, 1977a, on these aspects of verbal communication).

Another processing assumption that may have to be reevaluated concerns our decision to describe the two aspects of comprehension that the present model deals with—the coherence graph and the macrostructure—as if they were separate processes. Possible interactions have been neglected because each of these processes presents its own set of problems that are best studied in isolation, without the added complexity of interactive processes. The experimental results presented above are encouraging in this regard, in that they indicate that an experimental separation of these sub-processes may be feasible: It appears that recall of long tests after substantial delays, as well as summarization, mostly reflects the macro-operations; while for immediate recall of short paragraphs, the macroprocesses play only a minor role, and the recall pattern appears to be determined by more local considerations. In any case, trying to decompose the overly complex comprehension problem into subproblems that may be more accessible to experimental investigation seems to be a promising approach. At some later stage, however, it appears imperative to analyze theoretically the interaction between these two subsystems and, indeed, the other components

of a full-fledged comprehension model, such as the text-to-text-base translation processes and the organization of the text base in terms of facts.

There are a number of details of the model at its present stage of development that could not be seriously defended vis-à-vis some theoretical alternatives. For instance, we have assumed a limited-capacity, short-term buffer in the manner of Atkinson and Shiffrin (1968). It may be possible to achieve much the same effect in a quite different way, for example, by assuming a rapidly decaying, spreading activation network, after Collins and Loftus (1975) and Anderson (1977). Similarly, if one grants us the short-term buffer, we can do no more than guess at the strategy used to select propositions to be maintained in that buffer. That is not a serious problem, however, because it is open to empirical resolution with presently available methods, as outlined in this article. Indeed, the main value of the present model is that it is possible to formulate and test empirically various specific hypotheses about comprehension processes. Some rather promising initial results have been reported here as illustrations of the kind of questions that can be raised with the help of the model. Systematic investigations using these methods must, of course, involve many more texts and several different text types and will require an extended experimental effort. In addition, questions not raised here at all will have to be investigated, such as the possible separation of reproduction probabilities into storage and retrieval components.

To understand the contribution of this model, it is important to be clear about its limitations. The most obvious limitation of the model is that at both input and output it deals only with semantic representations, not with the text itself. Another restriction is perhaps even more serious. The model stops short of full comprehension because it does not deal with the organization of the propositional text base into facts. Processing proceeds only as far as the coherence graph and the macrostructure; the component that interprets clusters of propositions as facts, as outlined above, is as yet missing. General world knowledge organized in terms of frames must play a crucial role in this process, perhaps in the

manner of Charniak (1977). Note that this component would also provide a basis for a theory of inference, which is another missing link from our model of macro-operations.

In spite of these serious limitations, the present model promises to be quite useful, even at its present stage of development. As long as "comprehension" is viewed as one undifferentiated process, as complex as "thinking in general," it is simply impossible to formulate precise, researchable questions. The model opens up a wealth of interesting and significant research problems. We can ask more refined questions about text memory. What about the suggestion in the data reported above that gist and detailed information have different decay rates? Is the use of a short-term buffer in any way implicated in the persistent finding that phonological coding during reading facilitates comprehension (e.g., Perfetti & Lesgold, 1978)? One can begin to take a new look at individual differences in comprehension: Can they be characterized in terms of buffer size, selection strategies, or macroprocesses? Depending on the kind of problem a particular individual may have, very different educational remedies would be suggested! Similarly, we may soon be able to replace the traditional concept of readability with the question "What is readable for whom and why?" and to design texts and teaching methods in such a way that they are suited to the cognitive processing modes of particular target groups. Thus, the model to some extent makes up in promise for what it lacks in completeness.

References

- Aaronson, D., & Scarborough, H. S. Performance theories for sentence coding: Some quantitative models. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 277-303.
- Anderson, J. R. *Language, memory and thought*. Hillsdale, N.J.: Erlbaum, 1977.
- Anderson, R. C., & Pichert, J. W. Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 1-12.
- Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press, 1968.
- Bobrow, D., & Collins, A. (Eds.). *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.
- Chandler, P. J. *Subroutine STEPIT: An algorithm that finds the values of the parameters which minimize a given continuous function*. Bloomington: Indiana University, Quantum Chemistry Program Exchange, 1965.
- Charniak, E. A framed PAINTING: The representation of a common sense knowledge fragment. *Cognitive Science*, 1977, 1, 335-394.
- Clark, H. H. Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading*. Hillsdale, N.J.: Erlbaum, 1977.
- Colby, B. N. A partial grammar of Eskimo folktales. *American Anthropologist*, 1973, 75, 645-662.
- Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. *Psychological Review*, 1975, 82, 407-428.
- Halliday, M. A. K. Notes on transitivity and theme in English. *Journal of Linguistics*, 1967, 3, 199-244.
- Haviland, S. E., & Clark, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 512-521.
- Hayes, J. R., Waterman, D. A., & Robinson, C. S. Identifying the relevant aspects of a problem text. *Cognitive Science*, 1977, 1, 297-313.
- Heussenstam, F. K. Bumperstickers and the cops. *Transactions*, 1971, 8, 32-33.
- Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, 7, 194-227.
- Jackson, M. D., & McClelland, J. L. Sensory and cognitive determinants of reading speed. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, 556-574.
- Jarvella, R. J. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 409-416.
- King, D. R. W., & Greene, J. G. Invariance of inference times when information was presented in different linguistic formats. *Memory & Cognition*, 1974, 2, 233-235.
- Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.
- Kintsch, W. On comprehending stories. In M. A. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Erlbaum, 1977.
- Kintsch, W., & Greene, E. The role of culture specific schemata in the comprehension and recall of stories. *Discourse Processes*, 1978, 1, 1-13.
- 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. Summarizing stories after reading and listening. *Journal of Educational Psychology*, 1977, 69, 491-499.
- 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., Mandel, T. S., & Kozminsky, E. Summarizing scrambled stories. *Memory & Cognition*, 1977, 5, 547-552.
- Kintsch, W., & Monk, D. Storage of complex informa-

- tion in memory: Some implications of the speed with which inferences can be made. *Journal of Experimental Psychology*, 1972, 94, 25-32.
- Kintsch, W., & van Dijk, T. A. Comment on se rapelle et on résumé des histoires. *Langages*, 1975, 40, 98-116.
- Kintsch, W., & Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Memory: Processes and problems*. Hillsdale, N.J.: Erlbaum, 1978.
- LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, 6, 293-323.
- Labov, W. J., & Waletzky, J. Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts*. Seattle: Washington University Press, 1967.
- Longacre, R., & Levinsohn, S. Field analysis of discourse. In W. U. Dressler (Ed.), *Current trends in textlinguistics*. Berlin, West Germany: de Gruyter, 1977.
- Luria, A. R., & Tsvetkova, L. S. The mechanism of "dynamic aphasia." *Foundations of Language*, 1968, 4, 296-307.
- 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, The Netherlands: North Holland, 1975.
- Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.
- Neyman, J. Contribution to the theory of the test. In J. Neyman (Ed.), *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1949.
- Norman, D. A., & Rumelhart, D. E. *Explorations in cognition*. San Francisco: Freeman, 1975.
- Perfetti, C. A., & Goldman, S. R. Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 33-42.
- Perfetti, C. A., & Lesgold, A. M. Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. Weaver (Eds.), *Theory and practice in early reading*. Hillsdale, N.J.: Erlbaum, 1978.
- Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. Children's comprehension and memory for stories. *Journal of Experimental Child Psychology*, 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.
- Schank, R., & Abelson, R. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.
- Sgall, P., & Hajičová, W. E. Focus on focus. *The Prague Bulletin of Mathematical Linguistics*, 1977, 28, 5-54.
- Stein, N. L., & Glenn, C. F. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Multidisciplinary approaches to discourse processing*. Hillsdale, N.J.: Ablex, in press.
- Sticht, T. Application of the audread model to reading evaluation and instruction. In L. Resnick & P. Weaver (Eds.), *Theory and practice in early reading*. Hillsdale, N.J.: Erlbaum, in press.
- Turner, A., & Greene, E. The construction of a propositional text base. *JSAS Catalog of Selected Documents in Psychology*, in press.
- van Dijk, T. A. *Some aspects of text grammars*. The Hague, The Netherlands: Mouton, 1972.
- van Dijk, T. A. Text grammar and text logic. In J. S. Petöfi & H. Rieser (Eds.), *Studies in text grammar*. Dordrecht, The Netherlands: Reidel, 1973.
- van Dijk, T. A. Philosophy of action and theory of narrative. *Poetics*, 1976, 5, 287-338.
- van Dijk, T. A. Context and cognition: Knowledge frames and speech act comprehension. *Journal of Pragmatics*, 1977, 1, 211-232. (a)
- van Dijk, T. A. Macro-structures, knowledge frames, and discourse comprehension. In M. A. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Erlbaum, 1977. (b)
- van Dijk, T. A. Pragmatic macro-structures in discourse and cognition. In M. de Mey et al. (Eds.), *Communication and cognition*. Ghent, Belgium: University of Ghent, 1977. (c)
- van Dijk, T. A. *Text and context: Explorations in the semantics and pragmatics of discourse*. London, England: Longmans, 1977. (d)
- van Dijk, T. A. Recalling and summarizing complex discourse. In W. Burghardt & K. Hölker (Eds.), *Text processing*. Berlin, West Germany: de Gruyter, in press.
- van Dijk, T. A., & Kintsch, W. Cognitive psychology and discourse. In W. Dressler (Ed.), *Current trends in textlinguistics*. Berlin, West Germany: de Gruyter, 1977.

Received January 3, 1978 ■