46

The Construction–Integration Model of Text Comprehension and Its Implications for Instruction

Walter Kintsch

Comprehension: A Paradigm for Cognition

Understanding and *comprehension* are everyday terms, useful, but imprecise. We know what we mean when we say we understand a text, but understanding is difficult to define precisely: It is not necessary that we repeat the text verbatim, but we ought to be able to come up with the gist; it is not necessary that we think of every implication of what we have read, but we do not understand it if we miss the most obvious ones; it is not necessary that we answer every question that could be asked, but we cannot miss them all. In the laboratory as well as in the classroom, this problem is solved by fiat, operationally: We are willing to say that someone understands a text if he or she passes whatever test we have decided on: provide a summary, answer questions, verify inferences, and so forth. Not all of these operational definitions of understanding are equivalent, nor are they appropriate for all purposes. Much of the discussion in this piece aims at clarifying this situation, empirically, by showing what works where and for what purposes, and theoretically, by providing a framework that allows us to describe the different flavors of comprehension processes and outcomes.

There is, however, also a more technical use of the term *comprehension* that concerns us here. It is the sense in which comprehension is used in the phrase "comprehension as a paradigm for cognition." Cognition ranges from perception on the one hand to analytic thought on the other. Typically, the processes of perception and thinking are conceptualized in different ways. Perception is usually considered as some sort of constraint satisfaction process, where the organism must make sense of a wide variety of sensory inputs involving several modalities, such as solving a puzzle in which the pieces could be assembled in several different ways; the best way is the one that violates the least number of constraints. Thinking or problem solving, in contrast, is a matter of planning, of generating search spaces

and using means–end strategies to find a solution path. Reading comprehension shares aspects with both. On the one hand, one normally just reads and understands, much like we understand when we look at a visual scene, without elaborate planning and effortful problem solving. But when this normal process breaks down, the reader (or perceiver) becomes a problem solver who must figure out what it is he reads (or sees). Comprehension in this technical sense is automatic meaning construction via constraint satisfaction, without purposeful, conscious effort. Normal reading involves automatic comprehension, as well as conscious problem solving whenever the pieces of the puzzle do not fit together as they should.

The theory of text comprehension outlined in this piece is a comprehension model in the sense discussed here, but it leaves room for problem solving and planning when that becomes necessary to complement normal reading. This is a matter with considerable educational implications because instruction by its very nature pushes readers beyond what they already know and are comfortable with, requiring active, effortful, resource-demanding problem-solving activities that are difficult to maintain and direct.

Cognition and Representation

Theorists interested in text comprehension talk about the outcome of comprehension in terms of mental representations. Considered most broadly, in the present context, a mental representation is some change in the way the mind views the world as a result of reading a text, that is, some sort of trace of the text read, including indirect effects, cognitive as well as affective ones—perhaps a tendency to act in a certain way or to feel good or bad about something. There is little agreement about mental representations (or the lack thereof) among cognitive scientists at this point, and it would be impossible to do justice to the complex literature in a brief review. But there are a few points that are directly relevant to text comprehension, and that are not overly controversial.

The mind represents different aspects of the world. It is convenient to talk about these as different types or levels of representation. In a reading context, the levels that concern us most directly are perceptual representations, verbal representations, and semantic representations. Perceptual representations may be images of how the words looked on the page or how they sounded when spoken by a particular person. They also may be, however, images of the scene described by the text, constructed by the reader. Verbal or linguistic representations are about the words, sentences, and discourses themselves. Semantic representations refer to the ideas expressed by the words. Obviously, these levels are not cleanly separable. A word has perceptual characteristics as well as meaning, but when we talk about how a word is perceived and remembered, it is useful to keep these different aspects separate because they behave differently. Similar visual forms are confused with each other, as are similar phonemes, but words are more often confused on the basis of semantic similarity; decoding is strongly influenced by word length and word frequency, but semantic relations and conceptual structure are more important for comprehension. Hence, psychologists, as well as educators, do well to differentiate between the various levels of mental representations.

There is one more reason for the distinction among levels of representation. Theorists and model builders can deal quite well with verbal and semantic representations, but so far they have not developed the tools to deal effectively with imagery. Various systems are in use to represent the meaning of words. Feature systems are used widely; for instance, *bachelor* has the features *male* and *unmarried*, plus some others. Alternatively, word meanings are represented by their position in a semantic structure: *Shark* is defined as a member of the category *fish*, with special properties, such as *dangerous*. Or one can define word meanings by their position in a semantic space: *Lion* might be characterized by high values on the dimensions *size* and *ferocity*, whereas *mouse* would have low values. High-dimensional, abstract semantic spaces are especially effective for representing the meaning of words. Propositions are idea units, combining more than one word in a schematic form: *The hiker watches the elk with his binoculars* is a conceptual unit that relates, by means of the predicate *watches*, an agent, object, and an instrument in a meaningful, conventional way. Propositions thus allow the theorist to represent the meaning of sentences, independent of their syntactic structure (e.g., a sentence in passive or active voice would be represented by the same proposition). Furthermore, propositions can be combined to form representations of whole texts, as described in more detail below. The structure of these text representations is of great significance, because it allows the theorist to distinguish important ideas from mere detail, and it predicts how a text is comprehended and remembered.

Propositional structures are useful to represent the meaning of a text because they tend to mimic the properties of how people represent the meaning of a text. As yet we do not have comparable systems to represent mental images. Pictures will not do, for much the same reason that a text is not well represented by the actual words used: The picture does not make explicit the psychologically important aspects of an image. In the auditory domain, phonemic features capture quite well the salient aspects of how people perceive and remember the sounds of a language. However, visual feature systems have been only partially successful and have limited use, for example, in the form of geons (simple, primary shapes, such as bricks, cylinders, cones, and wedges that compose visual images and distinguish one image from another, as in a shoe vs. an ice cream cone) (Biederman, 1987). While propositions provide the theorist with a convenient and workable representation for the meaning of texts, at present there really is no language that we can use to represent the salient features of complex mental images. This deficiency is a major reason why much of the research on text comprehension has focused on the verbal aspects, neglecting the role of mental imagery for all its acknowledged significance. We shall, however, point out that significance wherever possible.

Levels of Text Representation

Texts consist of words, organized into sentences, paragraphs, and higher-order discourse units such as sections or chapters. The mental representation a reader forms thereof often is called the *surface-level memory*—the memory for the actual words and phrases of the text. Surface memory is typically short-lived, especially for instructional texts, where it does not matter much exactly how something is said (Sachs, 1967). Where that matters, as in a poem, joke, or argument, exact wording can be remembered very well, however (W. Kintsch & Bates, 1977).

For many purposes, we are not concerned with exact wording but with the message conveyed. Thus, it is useful to distinguish a semantic level of text representation—the ideas expressed by the text. We shall call this the *propositional level* of representation because propositions are one way of specifying what constitutes an idea in a text.

For present purposes[1] we define an *atomic proposition* as a linguistic unit consisting of a relational term (or predicate) and one or more arguments (which may be concepts or other propositions). Some examples of phrases and their corresponding atomic propositions are as follows:

(1) *Little boy* or *The boy is little* → [LITTLE, BOY]

(2) *The boy chopped the wood* → [CHOP, BOY, WOOD]

Note that this representation does not represent all information in a sentence (e.g., the past tense in (2), which is not important enough in many situations in which such propositional representations are used).

A *complex proposition* is a network of atomic propositions corresponding to a (simple) sentence. Propositions are linked in a network either because they are related referentially, as in (3), or because of propositional embedding (in (4) the arguments of the proposition are themselves atomic propositions).

(3) *The little boy chopped wood* → [CHOP, BOY, WOOD] —— [LITTLE, BOY]

(4) *Although the boy was little, he chopped the wood* →

[ALTHOUGH] —— [LITTLE, BOY]

[CHOP, BOY, WOOD]

Links may be based on other-than-referential overlap among propositions, for example, on the basis of a causal relationship, as in the following sentence:

(5) *The little boy was tired from chopping wood* →

[TIRED, BOY] —— [LITTLE, BOY]

[CHOP, BOY, WOOD]

This form of propositional representation is intentionally crude; its purpose is not to represent the meaning of a text in all its considerable complexity but to make it possible to count idea units in a text in a reasonably principled way (W. Kintsch, 1974; van Dijk & Kintsch, 1983). Both for the purpose of psychological research on text and instructional design, the number of idea units as defined here and their interrelationship are major variables of interest. Usually, we are not interested in how many words someone remembers but in how many and which ideas are remembered. What makes reading difficult is determined not only by sentence length and the familiarity of the words used but also by the number of ideas expressed, their coherence, and their structure (W. Kintsch, 1974). Propositional analysis, therefore, has become a valuable research tool (although it is not a teaching tool). Unfortunately, because it depends on hand coding, it is extremely laborious and not fully objective (a current guide is W. Kintsch, 1998, chap. 3.1.1).

The syntactic information in a sentence largely determines the structure of the propositional network. For instance, the main verb of a sentence is taken to form the superordinate proposition, and modifiers are subordinated to it, as in (3). However, there is more structure in a discourse than the sentence syntax. Discourses are organized globally, often according to conventional rhetorical formats. Thus, the simplest stories are of the form setting-complication-resolution; instructional texts may employ various structures such as a compare-and-contrast schema, or a generalization-plus-examples schema. To distinguish this discourse-level structure from the sentence-level structure, the terms *macrostructure* and *microstructure* are used. The microstructure of a text is the network of propositions that represents the meaning of the text. One can think of it as a translation from the actual words used into an idea-level format. The macrostructure is the global organization of these ideas into higher-order units. Thus, a story may have many propositions, linked in a complex network, but at the macrostructure level these propositions are grouped into the conventional sections: setting, complication, and resolution. However, a writer also could have chosen a different way of telling his story, for example, starting with the resolution and then filling in the setting and complication in the form of a flashback. That approach yields a very different macrostructure, while the microstructure might not be changed very much.

Microstructure and macrostructure together form the *textbase*, the semantic underpinning of a text. However, for purposes of psychological research on text comprehension, as well as for understanding educational practice, it is important to distinguish a further level of text representation, the *situation model*. The situation model represents the information provided by the text, independent of the particular manner in which it was expressed in the text, and integrated with background information from the reader's prior knowledge. What sort of situation model readers construct depends very much on their goals in reading the text, as well as the amount of relevant prior knowledge that they have. Thus, cooperative and attentive readers will more or less form the same textbase micro- and macrostructures, as invited by the author of the text. But depending on their interests, purposes, and background knowledge, they may form widely different situation models. In instruction, it is usually the situation model that the student forms from reading a text that is of interest; the teacher does not care whether the student can recite the text but whether the student understood it correctly and, for future use, was able to integrate the textual information with whatever background knowledge there was.

Situation models are not necessarily verbal. Texts are verbal, and textbases are propositional structures, but to model the situation described by a text, people often resort to imagery. Mental images of maps, diagrams, and pictures are integrated with verbal information in ways not well understood by researchers. Individual preferences in this regard further limit the ability to predict just what sort of a situation model a reader will form from a text.

It is important to ask not only whether a good, correct situation model has been formed by a reader reading a text, but also whether this new model has been integrated with the reader's prior knowledge. It is quite possible that readers may construct adequate textbases but fail to link them with other relevant portions of their prior knowledge. The result is encapsulated knowledge. If readers are reminded of the text from which they have acquired this knowledge, they can remember it and successfully use this knowledge, but it is not part of their generally available knowledge base. Encapsulated knowledge can be retrieved only via the specific episodic text memory; it is not available on occasions when such knowledge may be useful but the episodic retrieval cues are lacking. Thus, students can do their calculus problems at the ends of the chapters in their textbooks, and even on final exams, but have no idea what to do when they are supposed to use their knowledge in an engineering class. To make knowledge acquired from texts usable in novel situations, it must be actively linked to semantic retrieval cues, which is not an automatic process but one that requires strategic action and effort on the part of the reader/learner.

Example: Levels of Representation

"Connected" is a story of about 2,500 words written with the purpose of teaching novice students some basic facts about electricity that are embedded in the story in the form of explanations provided by a father to his daughter, who is trying to solve a puzzle requiring knowledge of these facts. The story has four subheadings: "An important event," "Life on the farm," "How does electricity work?" and "Solving the mystery."

The surface memory for this text refers to whatever sentences and sentence fragments from this text are still available in the reader's memory, and it need not concern us further here.

The macrostructure of the text is shown in Table 1. It is basically a high-level summary, organized according to the classical story schema. It only roughly corresponds to the subheadings of the actual text: The setting comprises the first and part of the second section, the complication corresponds to part of the second and the third sections, and the resolution matches the final section.

An example of the microstructure for this text is given for one brief paragraph in Table 2. The first sentence is represented as two complex propositions, C1 and C2, each consisting of three atomic propositions. C1 and C2 are linked by P4, i.e., the sentence connective *while*. The third and fourth sentences of the paragraph are each represented by a complex proposition, C3 and C4. Note that anaphoric inferences are necessary here: the *she* of the text has to be identified as *Katie*. To understand, a further inference is required: The *appliances* in the last sentence must be identified with the *electric iron*, the *electric lamp*, and the *electric sewing machine* mentioned earlier. For an adult reader, this is an automatic inference, made unconsciously and effortlessly. For a child, however, who does not really know what an appliance is, this may be a major stumbling block, requiring the reader to regress and figure out that the *appliances* are the *lamp* (which makes *light*), the *iron* (which makes *heat*), and the *sewing machine* (whose parts *moved*). What is necessary here is a conscious, strategic process of meaning construction, which is effortful and resource demanding. The reader who avoids this effort still can form a coherent textbase—Katie realizes that appliances make heat and so on—but will be unable to construct an adequate situation model without knowing what *appliances* refers to.

TABLE 1
The Macrostructure of the Story "Connected"

Setting
Location: on a farm
Time: old days
Actors: Katie, Tom, and their parents

Electricity is coming to town.
The children wonder what sort of appliance their parents are going to buy.

Complication
Their father asks them to guess what electricity-using appliance they will get first.
Katie finds out how electricity works and what it is used for.
She finds out that the appliance is not to produce either heat or motion.

Resolution
Because there are two wires on the electric line being installed, the first appliance their parents buy will be a telephone.

TABLE 2
The Microstructure for One Paragraph

In town, her father filled the Model T's gas tank, while Katie bought a sewing machine belt and browsed in the general store. She saw an electric iron, electric lamps, and a sewing machine that no one had to pedal. She realized there were ***appliances*** *that made heat and light and those that moved.*

C1	P1	[IN, TOWN, P2]
	P2	[FILL, FATHER, GAS-TANK]
	P3	[HAS-PART, MODEL-T, GAS-TANK]
	P4	[WHILE, P2, P5, P7]
C2	P5	[BUY, KATIE, BELT]
	P6	[HAS-PART, SEWING-MACHINE, BELT]
	P7	[BROWSE, KATIE, GENERAL-STORE]
C3	P8	[SEE, **KATIE**, IRON, LAMPS, SEWING-MACHINE]
	P9	[ELECTRIC, IRON]
	P10	[ELECTRIC, LAMPS]
	P11	[NOT-HAVE, SEWING-MACHINE, PEDAL]
C4	P12	[REALIZE, **KATIE**, P13, P14]
	P13	[MAKE, APPLIANCE, HEAT, LIGHT]
	P14	[MOVE, APPLIANCE]
	INF	[IS, **APPLIANCES**, P9, P10, P11]

For explanation, see page 1276.
C = a complex proposition; P = an atomic proposition.

What sort of situation model might a student construct upon reading this story? The student's reading goal is to learn about electricity. Hence, the situation model we are interested in concerns what the student has learned about electricity; the story is merely there to keep up the students' interest. Skillfully interwoven into our story is a puzzle, the "mystery" Katie must solve for which one needs to know certain elementary facts about electricity. The students are not faced with a list of dry facts about electricity but with information that is significant for the puzzle they—and Katie—are trying to solve. Table 3 lists these facts as the situation model a successful reader will form and link to whatever he or she already knows about electricity.

To construct Table 3, hypothetical prior knowledge for a typical reader has been assumed; any real reader may not know exactly what is listed. What is important is that the reader retrieve such pieces of prior knowledge at the right moment when reading this story so that they can become associated with the new information provided by the text. Thus, suppose a reader already knows that electricity is needed for ironing; now he or she learns that the electric energy generates heat in the process of ironing, and if this new bit of information is linked with what is already known, it successfully becomes a part of the reader's knowledge base, not just an item of information remembered in the context of that particular text.

TABLE 3
The Situation Model

Prior Knowledge	Information Provided by the Text
Electricity needs **wires**.	Electricity comes to you via **wires** (you do not have **to get it like wood to burn in the stove**).
You have to **bring wood in for the stove.**	Electricity is generated from coal or water. Static electricity produces sparks and lightning. Electric current is a form of energy (like water power).
Electricity is needed for **lamps**, **ironing**, **sewing machines**, and **telephones**.	Electric energy is used to make **light** (by heating up the filament in a bulb). make heat (such as for **ironing**). make motion (such as in a **sewing machine** or record player). talk on the **telephone** (which needs extra wire and was invented by Bell).

Links are formed between corresponding items printed in bold.

Surface structure, textbase, and situation model are levels of the mental representation of texts. We next turn to the question of how these representations are constructed.

The Process of Comprehension: Construction and Integration

Most of the research on reading deals with the decoding problem: How do readers translate the written text into words and sentences? In other words, how is the surface representation generated from a written text? This is, of course, an extremely important question with complex answers, but it is not the question that will be addressed here. Instead, we shall assume this level of representation as given and look at the formation of the textbase and the situation model.

Microstructure

Given a text—a structured string of sentences—how are the corresponding idea units derived, and how are they organized? For the most part, the language provides good cues as to the underlying ideas: *The goat ate the grass* unproblematically translates into [EAT, GOAT, GRASS]. However, language is full of ambiguities. We understand both of the following sentences:

(6) *The grade was too steep.*

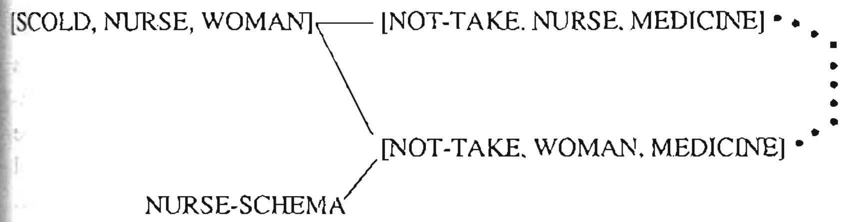
(7) *His grade was an A.*

And we know who *she* and *he* are in these two sentences:

(8) *The nurse scolded the woman because she had not taken her medicine.*

(9) *The hiker saw the grizzly bear. He was afraid.*

There are two kinds of explanations of how people deal with such ambiguities, top-down theories and bottom-up theories. According to the top-down view, a schema filters out incorrect interpretations: We know we are talking about a hill, or a student, and hence assign the right meaning to *grade*; the nurse–patient schema dictates the referent for *she*; and the grizzly bear schema specifies who has to be afraid. Schema theory is very powerful (e.g., Schank & Abelson, 1977), and schema effects in perception and comprehension are well documented. Nevertheless, schema-as-filter theories of comprehension cannot fully account for comprehension processes and have been replaced by theories that assign a more decisive role to bottom-up processes, such as the Construction–Integration (CI) model (W. Kintsch, 1988, 1998). Instead of trying to construct only the correct meaning of a sentence, the CI model generates several plausible meanings in parallel and only later, when a rich context is available, sorts out which construction is the right one. This sorting out is done by means of an integration or constraint satisfaction process that suppresses those constructions that do not fit in well with the context and strengthens those that do. Specifically, activation is spread around in the propositional network that has been constructed, including the contradictory elements; the activation eventually settles on those nodes of the network that hang together, while outliers and isolated nodes become deactivated. Thus, in (6) and (7) propositions will be constructed initially involving both meanings of *grade*, but the incorrect meaning will become deactivated during the integration phase. For the anaphora identification in (8), the construction process yields



where the dotted line indicates an inhibitory link. In the integration process, the correct proposition will win out because it is connected to prior knowledge about nurses and patients (here labeled the NURSE-SCHEMA). Thus, schemas play a role in the CI model, too, not as filters that control construction but as context that influences the integration process. The inference in (9) is handled similarly: In the construction phase, the model is not sure whether the *hiker* or the *bear* is afraid, but prior knowledge settles that question during the integration phase.

Thus, the CI model uses a bottom-up construction phase in which contradictory assumptions are explored, resulting in an incoherent network that needs to be cleaned up in the integration phase. The computational advantage of such a dual process is that the construction rules do not have to be very smart, because errors can be corrected in the integration phase. Psychological data that suggest that human comprehension processes employ a similar scheme are discussed in a subsequent section on word identification.

To illustrate the construction of a microstructure, let us return to the "Connected" story discussed earlier. The list of propositions in Table 2 corresponds to the network shown in Figure 1. The links in Figure 1 are based on referential overlap between propositions. Two obligatory inferences are required to identify the pronouns for P8 and P12.

The final activation values for the network in Figure 1, once the process of spreading activation has stabilized, are shown in Figure 2. Figure 2 implies that after reading this paragraph, the strongest information in memory should be that Katie bought a belt, browsed in the general store, and saw an electric iron, electric lamps, and a sewing machine. On a recall test, those should be the items most frequently recalled. A large number of studies have borne out such recall predictions (e.g., W. Kintsch, 1974).

Also shown in Figure 2 are the strength values obtained if the reader makes the optional inference [IS-APPLIANCE, IRON, LAMP, SEWING-MACHINE]. This inference changes the picture a great deal by emphasizing the relationship between the (complex) propositions corresponding to the last two sentences of the text. It will be remembered that this was an instructional text supposed to teach about electricity. Note that without this "deep" processing (the inference about appliances) the present paragraph would not contribute much to the goal of learning physics.

Macrostructure

Generally (except for the case of very brief texts), understanding a text requires formulating a mental representation of its macrostructure. Just what role a proposition plays in a text depends on its function in the overall structure: It may be part of the gist of an essay, or it may be an expendable detail; it may be a crucial link in the causal chain of a story, or it may be irrelevant to the main story line. To capture this kind of intuition, van Dijk (1980) has introduced the concept of a macrostructure. The macrostructure of a text consists of those propositions that are globally relevant, that form its gist in everyday language. Macrostructures are frequently but not necessarily schematic; that is, they are based on conventional rhetorical forms. Thus, narratives have a conventional structure in our culture; essays may be in the form of arguments, or definitions-plus-illustrations, and so on (see van Dijk & Kintsch, 1983, chap. 2.9, for a detailed discussion). Van Dijk (1980) has enumerated three rules that describe the formation

FIGURE 1
The CI Network for a Paragraph From the "Connected" Story (Corresponds to the Proposition List in Table 2)

IN-TOWN	c0	0.1806	(P1)
FILL-GASTANK	c0	0.5027	(P2)
OF-MODEL-T	c0	0.1806	(P3)
WHILE	c0	0.8604	(P4)
BUY-BELT	c0	0.2852	(P5)
FOR-SEWING-MACHINE	c0	0.3566	(P6)
BROWSE-STORE	c0	0.9063	(P7)
SEE	c0	1.0000	(P8)
ELECTRIC-IRON	c0	0.4474	(P9)
ELECTRIC-LAMP	c0	0.4474	(P10)
SEWING-MACH-WITHOUT	c0	0.4474	(P11)
REALIZE	c0	0.8809	(P12)
APPLIANCE-MAKE-LIGHT	c0	0.4474	(P13)
APPLIANCE-MAKE-MOVE	c0	0.4474	(P14)
FATHER	c0	0.1806	(V1)
KATIE	c0	0.7863	(V2)
IS-APPLIANCE	c0	1.0000	(S1)

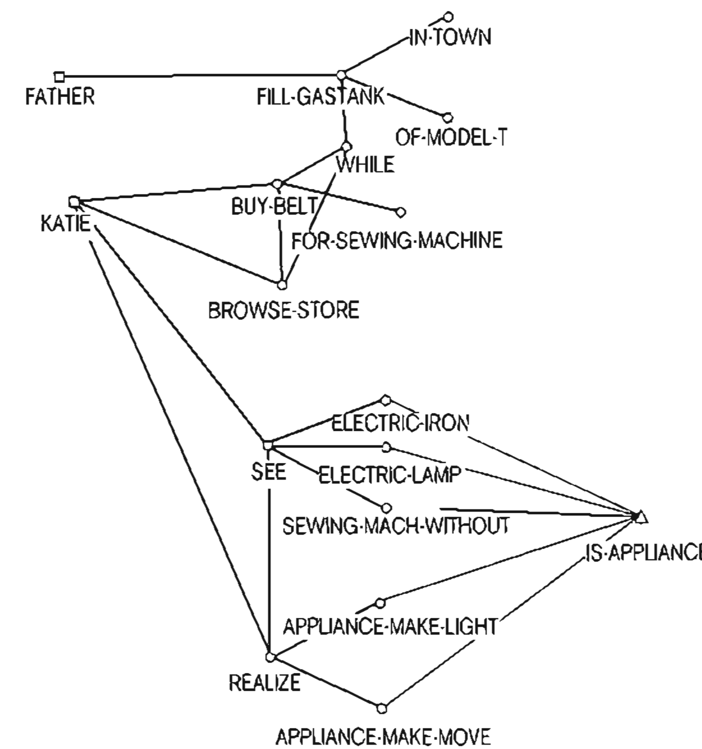
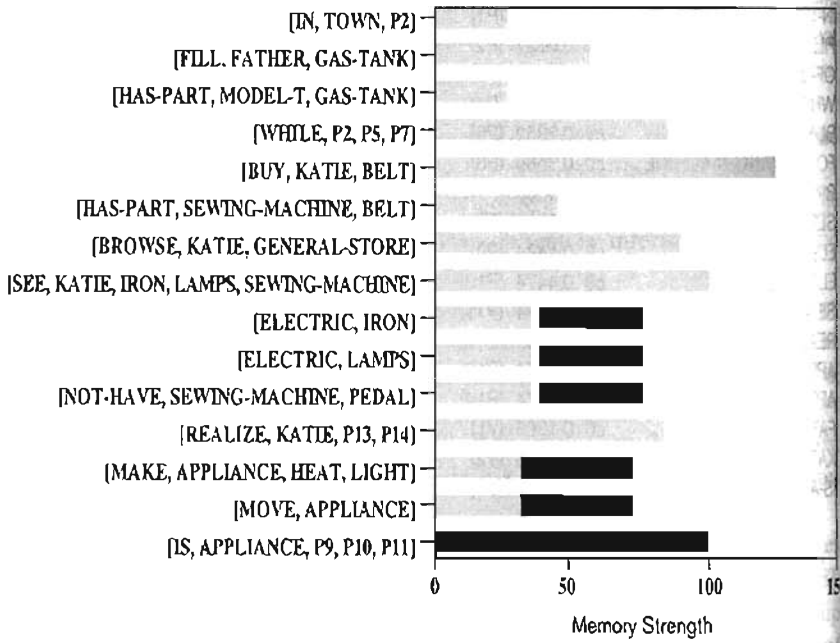


FIGURE 2
The Result of the Integration Process for the Network in Figure 1, Without the Inference (IS-APPLIANCE) and With It (Darker Bars)



of macrostructures: (1) *Selection* of macrorelevant propositions (and correspondingly the deletion of propositions that are not macrorelevant), (2) *Generalization*, that is, substitution of a superordinate proposition for subordinate propositions; and (3) *Construction*, the substitution of a general proposition describing a whole sequence of interrelated propositions. Given a text and a set of macropropositions, these rules can be used to show how the macropropositions were derived from the text. However, these rules are post-hoc: They describe how macropropositions were derived after the fact, but they are not rules that allow us to generate macropropositions from a text. They do not tell us what is to be deleted or what is to be generalized. In order to use these rules, one must already know what is macrorelevant, what can be subsumed under a construction, and so on. In other words, the macrorules are incomplete because they do not include the conditions for their application. This shortcoming has seriously limited the modeling of macrostructures, which is unfortunate because macrostructures play such an important role in comprehension (e.g., W. Kintsch & van Dijk, 1978).

A logical analysis of the relations linking linguistic units that overcomes some of the limitations of macrorules has been suggested by Le (2002), who distinguishes three types of relations among text units: (1) coordination (either in the form of elaboration or parallelism), (2) subordination, and (3) superordination. After one specifies the relations among text units (sentences or complex propositions), hierarchical structures at levels higher than the sentence can be generated that allow the identification of macropropositions. To illustrate Le's procedure, consider the brief paragraph analyzed in Table 2 that consists of four complex propositions, C1–C4. As shown in Table 4, C1 is subordinated to C2; C2 and C3 are coordinated, C3 being an elaboration of C2. C4 is logically superordinated to C3, because it expresses a generalization based on C3. Thus, Le's analysis identifies C4—the complex proposition at the highest level in the paragraph hierarchy—as the macroproposition for that paragraph.

A different approach to the generation of macrostructures has been taken by W. Kintsch (2002). It is not based on a logical analysis of the relations among text units, but rather on the centrality of the content of the (complex) propositions. Latent Semantic Analysis (LSA; discussed in a later section) allows one to measure the similarity of the content of sentences. The sentence in a paragraph that is most similar to all the other sentences in that paragraph is a good candidate for a macroproposition because it is the most central one. In Table 4, C3 correlates most strongly with the other sentences, as measured by LSA, and hence should be considered as the macroproposition for this paragraph. Note that this is a different result than the one obtained from Le's logical analysis. There is no reason why two so totally different methods should yield identical results; large-scale empirical tests of which predictions correspond best with human judgments have not yet been reported. Note also that in terms of the activation values for complex propositions as shown in Figure 2, the most strongly activated complex

TABLE 4
Determining the Macroproposition for a Paragraph

C1 *In town, her father filled the Model T's gas tank,*
C2 *while Katie bought a sewing machine belt and browsed in the general store.*
C3 *She saw an electric iron, electric lamps, and a sewing machine that no one had to pedal.*
C4 *She realized there were appliances that made heat and light and those that moved.*

$$\begin{tikzcd} & C4 \\ C2 \arrow[r] \arrow[d] & C3 \arrow[u] \\ C1 & \end{tikzcd}$$

proposition is C2. (Activation values for complex propositions are obtained by adding the activation values of their constituent atomic propositions.)

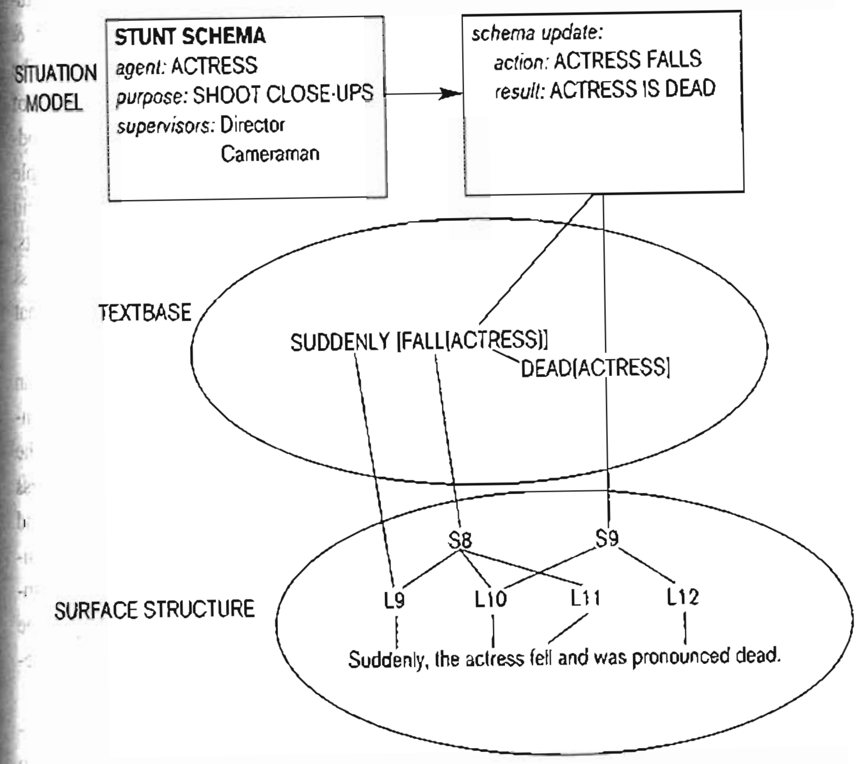
It has long been known that gist-level—that is, macrostructure—processes play a decisive role in the comprehension and memory of long texts. That much was shown by W. Kintsch and van Dijk in their 1978 paper. Modeling the generation of macrostructures, however, is still in its infancy, as the earlier discussion illustrates. Worse, there are basic limitations to the approaches of Le (2002) and W. Kintsch (2002): Both models can only select from the propositions in a text, whereas macropropositions often must be constructed by the reader. Macropropositions frequently are inferences that are not stated explicitly in the text. Computational models specifying how new macropropositions are generated do not yet exist. This is an important area for future research, as is the research on the formation of situation models, which is in a similarly underdeveloped state.

Situation Models

The problems faced by the researcher trying to model the formation of situation models are formidable. Textbases at the micro- and macrolevel are tightly constrained by the nature of the text, which a faithful reader must respect. The text, however, is only one factor in the situation model: The reader's goals, interests, beliefs, and prior knowledge also must be taken into account. Generally, these are only incompletely known. Furthermore, even the form that a situation model takes is not fully constrained: Situation models may be imagery based, in which case the propositional formalism currently used by most models fails us. Nevertheless, in well-defined contexts, modeling situation models is quite feasible and will surely be the focus of research on text comprehension in the next decade.[2]

How one might approach this task has been demonstrated by Schmalhofer, McDaniel, and Keefe (2002). The CI model simulates the construction of a textbase: A network of propositions derived from the text is constructed and integrated via a spreading activation constraint satisfaction process. Schmalhofer et al. added to the propositional network two other networks: (1) a surface level, where the nodes are linguistic structures and words, and (2) a situation representation, where the nodes are schemas. Nodes are interconnected at each level, but importantly, there are also links between levels, so that a sentence in the surface structure is connected to the corresponding proposition in the textbase, which in turn is connected to the appropriate schema at the situation model level. Schmalhofer et al. illustrate their model with an example that is reproduced here in simplified form in Figure 3. The text is a story about a movie stunt that results in a fatal accident. For the surface level of analysis, one sentence is shown, with word units L9 to L12 and syntactic units S8 and S9; of course, all this is part of a much larger network with rich interconnections not shown here. The units at the surface level are connected not only to each other but also to the propositional units at the textbase level. The propositions of the textbase are linked, in

FIGURE 3
Surface Structure, Textbase, and Situation Model



Adapted from Schmalhofer et al. (2002).
L = word units; S = syntactic units.

turn, to the situation model units, which here are schemas. The STUNT-SCHEMA has been partly filled in with information from previous portions of the text, but it is updated now with current information from the sentence being processed: An action and a result slot are filled in. When activation is spread in such a triple network, it is the structure present at each level of analysis that determines which nodes get activated, and complex interactions between levels also occur. This has important consequences, especially for the maintenance of inferences, as Schmalhofer et al. show. A model such as this explains how inferences at the situation model level can become integral parts of text memory, solidly and permanently anchored in the text structure.

The approach to modeling situation models pioneered by Schmalhofer et al. is a very promising one. Still, there are some limitations: The researchers selected their story in such a way that schema units were appropriate to represent the situation model. Not all situation models can be represented by schemas, however, and a more general approach is required. What needs to be represented in a situation model is at least partially understood (see discussion in Zwaan & Radvansky, 1998). Louwerse (2002) has developed a formal language to describe situations, with an emphasis on the coherence among units. His description of different types of coherence relations could potentially be incorporated into models such as that of Schmalhofer et al., extending their generality beyond simple script-based stories. The goal of discourse research should be to develop and evaluate situation models for complex narratives, and especially declarative texts, such as chapters from a science text, at the same level of detail and explicitness as Schmalhofer et al. have done for simple stories. If we understand better what students have to do, we shall be better able to guide and help them.

To summarize the research on the processes of text comprehension, we can say that we have a good understanding of how people go from the words and sentences of a text to the underlying ideas and how the text structure determines the organization of these ideas into a coherent textbase, at least at the local level. Less is known about global organization, or how readers form macrostructures, and even less is known about how situation models are constructed through the interplay between texts, background knowledge, and reader goals. However, promising beginnings have been made in these areas, and rapid progress can be expected now that reading researchers are placing more emphasis on comprehension rather than on the decoding aspects of reading.

The first part of this chapter has described a general theory of comprehension. In the next sections, the focus will be on the application of that theory to important research topics in the area of discourse comprehension: how words are identified in a discourse context, the representation of knowledge, the construction of macrostructures and situation models, the role of inferences and working memory, and the contrast between text memory and learning from text. Of particular interest are the implications of these research results for instruction, which will be emphasized throughout this discussion.

Word Identification

A great deal of research has gone into determining how the letter shapes on a page are turned into meaningful words. The results of this work will not be reviewed here because they have been discussed in other chapters of this volume (see Juel & Minden-Cupp, #13; Ehri & McCormick, #14; Kuhn & Stahl, #16; Nagy & Scott, #19). Instead, a body of research will be introduced here that complements this research in that it is concerned with the question of how readers arrive at the correct sense or meaning of a word[3] when they encounter it in a discourse context. To give a concrete example of what the issue is here, consider the following sentences:

(10) *A beautiful sight in downtown Denver is the <u>mint</u>.*

(11) *A fragrant tea is made with <u>mint</u>.*

How do we know that *mint* is a building in the first sentence but the leaves of a plant in the second? *Mint* is a homonym in English, that is, a word with more than one meaning, and readers obviously and effortlessly find the right meaning when they read (10) and (11). Similarly, when words with only a single meaning are used in different senses, readers readily perceive what is meant:

(12) *The fox <u>ran</u> faster than the hedgehog.*

(13) *The chancellor's decree <u>ran</u> into strong opposition.*

One explanation of how readers identify word meanings in context assumes that all word meanings and word senses are listed in a mental lexicon and that readers must select the right meaning or sense for the given context. There are at least two ways in which this selection could occur:

1. The schema acts as a filter. Suppose that each word meaning/sense in the mental lexicon is associated with a specific context. Thus, *mint* in (10) is associated with *a building-in-which-coins-are-manufactured* schema; reading (10) activates this schema, and the schema selects the proper sense of *mint* from the list of available senses. This is a top-down model, where the schema acts like a filter, admitting only the schema-relevant meaning and not admitting irrelevant meanings. Models of this type have been proposed by, among others, Schank and Abelson (1977).
2. The context suppresses inappropriate meanings. According to this model, all meanings/senses of a word are activated when reading it, but inappropriate meanings are suppressed by the context because they do not fit the contextual constraints. When reading (10), all versions of *mint* in the mental lexicon would be activated initially, but only one—the one associated with *building*, *downtown*, and *Denver*—would be consistent with the sentence context and would survive. Models of this type have been proposed by, among others, Swinney (1979).

Fortunately, it is possible to decide among these alternatives experimentally. Till, Mross, and Kintsch (1988) have reported a relevant experiment using the "lexical decision" method. In this experiment, participants read sentences such as (10) and were then asked to decide as quickly as possible whether a briefly presented string of letters was an English word or not. Four types of test items were used (each participant saw only one of these):

1. a nonword string (e.g., *baher*) for which the correct response was "no";
2. an associate of the target word that was contextually appropriate (e.g., *money*);
3. an associate of the target word that was contextually inappropriate (e.g., *tea*)
4. an unrelated control word (e.g., *baker*)

The correct response for the last three items was "yes," but interesting differences in response speed were observed. When the test item was presented immediately after the sentence, response times for associated items were significantly shorter than response times to unrelated control items, whether or not the association was contextually appropriate. That is, *mint* in (10) primed both *money* and *tea*. When the test item was presented with a 350 millisecond (msec) delay after the sentence, the response time for the contextually appropriate associate was shorter than the response time for either the control word or the inappropriate associate. That is, 350 msec after reading (10), only *money* was primed, not *tea*.

The Till et al. data clearly contradict the schema-as-filter model and support a model that posits a bottom-up activation of all word meanings, followed by a contextual constraint satisfaction process that deactivates inappropriate meaning. Indeed, these data were one of the original inspirations for the CI model (W. Kintsch, 1988). Today there exists a very large and complex literature on this subject, which cannot be reviewed here (but see, e.g., Rayner, Pacht, & Duffy, 1994). Results depend on various boundary conditions, but on the whole they effectively rule out the schema-as-filter model. It appears that, generally, multiple meanings and senses of a word are activated initially but that context-inappropriate meanings and senses are suppressed rapidly.

It is difficult to imagine, however, how such a meaning selection model could work. Just what are the cues that allow the selection of the right meaning or sense among so many alternatives? Furthermore, just what are the alternatives in the mental lexicon? How do we decide how many meanings or senses a word has? People learn to use words in ever-novel ways. Can a mental lexicon in which every use must somehow be explicitly defined do justice to this complexity? What if the different word meanings and senses are not predefined in a mental lexicon but emerge in context? How could such a generative lexicon be constructed? One attempt to do so invokes the idea of semantic elements that can be combined to form all meanings, much like the 100+ chemical elements can be combined to form all the manifold substances in the universe. This approach has not been successful, however, because nobody has been able to come up with a principled list of semantic elements or the rule system that would allow us to construct all meanings from the combination of these elements. An alternative approach that appears promising to achievement of the goal of a generative lexicon is based on some recent developments in statistical semantics. It will be briefly described in the next section.

Knowledge Representation: Latent Semantic Analysis

There exists no complete model of how the totality of human knowledge—including perceptual knowledge, verbal knowledge, action knowledge, and emotional knowledge—is represented in the mind and how it is organized. In the study of reading we are primarily concerned with verbal knowledge. Recently, a model of how human verbal knowledge is represented has been developed which is of considerable use for modeling the use of knowledge in comprehension. The model is called Latent Semantic Analysis (LSA) and provides a good account of the associative basis of human verbal knowledge, although not of analytic, logical thought. Essentially, it allows us to measure the semantic similarity between words, sentences, and whole texts in an objective and automatic way that agrees quite well with human judgments. Full descriptions of LSA are available in Landauer (1998), Landauer and Dumais (1997), and Landauer, Foltz, and Laham (1998).

LSA is a machine learning method that lets a computer induce a semantic structure merely from reading a large amount of text. The computer keeps track of how words are used in many different documents and then uses a mathematical procedure to extract the essential semantic relations among words and texts from this mass of information. The result is a high-dimensional semantic space in which meanings are represented mathematically by vectors.

A good way to think about LSA is in analogy to maps. The data on which maps are based are measurements of various features of the earth: distances, altitudes, and so on. Once we have drawn a map, these features are represented in two dimensions in a conventional way: An inch on the map corresponds to so-and-so many miles on the earth, contour lines indicate altitudes, blue indicates a lake or a sea, and so on. The map is useful because it goes beyond the data: It allows us to compute numerous relations among places that were never directly measured. LSA, analogously, provides a map of meanings. This map, too, is based on a certain set of observations, but it goes beyond the data on which it was based: It allows us to compare the meanings of any points in our map. Meaning relations, however, are much more complex than the relations among points on the surface of the earth and require many more dimensions to represent. In fact, LSA needs about 300 dimensions to satisfactorily map meaning. Thus, meanings are represented not by points but by vectors in a 300-dimensional space. Meaning in LSA is a sequence of 300 numbers, which is meaningful only in comparison to other such meaning sequences. The usual way to measure the distance between vectors is by computing their cosine in the high-dimensional geometric space. A cosine of 1 means identity, and a cosine of 0 means unrelated—much like a correlation coefficient. Here are some examples of cosines between words:

(14) *cop – cops*	.78
cop – policeman	.68
buy – bought	.66
high – low	.79
he – his	.97
blackbird – bird	.46
blackbird – black	.04

These examples illustrate both the strength and the limitations of LSA. It is a strength that LSA groups together words that are related in meaning. Note that antonyms are very close semantically, differing only in one—crucial—respect; LSA captures this relation but lacks the analytic power to distinguish between opposites. LSA also neglects syntax, with the consequence that *he* and *his* mean just about the same thing to LSA. Another limitation is illustrated with the *blackbird* example. LSA knows only what is explicit in the texts used to teach it; thus, *blackbird* is a *bird* for LSA, but it is not *black*. (When *blackbird* is used in the training text, it is used in a *bird* context, so LSA makes the necessary inference, but texts do not mention the color of *blackbirds*. We can decompose the word, but LSA cannot.)

The real power of LSA arises because it allows us to compare the meaning of sentences and, indeed, whole texts, as readily as words. Consider the following sentence triplet:

(15) *The airplane rescued the hiker.*

cos = .43 *cos* = .09

The climber was evacuated by helicopter. *The hiker saw the bear.*

The sentence pair that is similar in meaning according to our intuition is close in the semantic space (the two sentences have a high cosine), whereas the sentences that appear unrelated to human intuitions are distant in the LSA space (low cosine). Note that this happens in spite of the fact that the semantically related sentence pair does not share any content word, whereas the unrelated pair does. This ability to measure the similarity of texts is the basis for many successful applications of LSA, for example, for grading essays (Landauer, Laham, & Foltz, 2000) and for helping students to write better summaries (E. Kintsch et al., 2000).

Macrostructures and Summaries

Macrostructures are mental representations of text at a global level. They may simply mirror the structure of the text from which they were derived, or they may reflect, to varying degrees, the comprehender's own prior knowledge structure that has been imposed on the text in the creation of a situation model.

Macrostructures as envisaged by van Dijk (1980) and discussed in van Dijk and Kintsch (1983) are hierarchies of propositions. Macropropositions put into words are summary statements at different levels of generality. They subsume what the different sections of a text are about. They are derived from the text by the operations of selection, generalization, and construction, but propositional macrostructures cannot be computed automatically from a text. The macrorules merely help us to explain what can be done, but they are not algorithms or computational procedures that generate macropropositions from a text automatically. A computationally more feasible—but in other ways more limited—alternative for the representation of macrostructures is provided by LSA. Instead of representing the meaning of a sentence by a proposition, the meaning can be represented as a vector in an existing high-dimensional semantic space. For some purposes, such a representation is all that is needed. For example, one can compare new texts, such as summaries students write, with these macrovectors; one can compute the importance or typicality of sentences from the text, and so on.

For other purposes, verbal statements corresponding to macropropositions are needed. W. Kintsch (2002) has described how LSA can be used to select topic sentences from a text and to generate a summary by concatenating these topic sentences. There is more to a summary than just selecting topic sentences, but it is instructive to see what can be achieved in that way—and what is still missing. The text analyzed in W. Kintsch is a chapter titled "Wind Energy," taken from a junior high school science textbook. It is 960 words long and divided by its author into six sections, each with its own subtitles. Thus, the author indicates the intended macrostructure and even provides appropriate macropropositions, in the form of six subtitles. Macrorules can be used to explain where these subtitles come from. Consider the following paragraph (the second subsection of the chapter):

(16) *The history of windmills*

Since ancient times, people have harnessed the wind's energy. Over 5,000 years ago, the ancient Egyptians used the wind to sail ships on the Nile River. Later, people built windmills to grind wheat and other grains. The early windmills looked like paddle wheels. Centuries later, the people in Holland improved the windmill. They gave it propeller-type blades. Holland is still famous for its windmills. In this country, the colonists used windmills to grind wheat and corn, to pump water, and to cut wood at sawmills. Today people still sometimes use *windmills to grind grain and pump water, but they also use new wind machines to make electricity.*

The macrorule of construction can be used to compress sentences 2–4 into

People used wind energy in Egypt.

Similarly, the other sentences of the paragraph can be reduced to

People used windmills in Holland.

People used windmills in the colonies.

People use windmills today.

These sentences can be transformed by the macrorule of generalization into

(17) *People used windmills throughout history.*

or

(18) *The history of windmills.*[4]

Thus, macrorules allow us to postdict, or explain, what the author did. But the application of these rules depends on our intuitions about the text and our knowledge about it. By themselves, these rules cannot compute anything.

LSA provides a computational mechanism that can compute macrostructures of a kind. For instance, we can compute a vector in LSA space that is the centroid of all the words in paragraph (16). Such a vector may seem to be totally useless—it is, after all, a list of 300 uninterpretable numbers—but that is not so. It can be quite useful, for instance to decide how appropriate a proposed subtitle is. The cosine between the paragraph vector and the proposed subtitle is a measure of how close the subtitle is to the paragraph as a whole. For instance, (17) and (18) have rather similar cosines with the paragraph—.39 and .48, respectively, high enough to indicate that they are both acceptable summary statements. But suppose we had chosen an ill-considered subtitle for the paragraph, such as "Holland is still famous," or something totally inappropriate such as "Rain douses forest fires." The cosine measure would have allowed us to reject these choices (the cosine is .26 in the first case and only .05 in the second, both much lower than the cosines for (17) and (18)).

There are other uses for vector representation of a macrostructure, too. For instance, we can compute how closely related the sections of a text are to each other. This kind of information can be of interest in various ways. If two sections of a text are very closely related, one might consider combining them. Or if two similar sections are separated by a dissimilar one in the text, one might consider reordering the sections of the text. We also can obtain a measure of how important a section is to the overall text. One way to do this is to compute the cosine between the whole text and each section.

To generate the full range of macropropositions is beyond the scope of LSA; operations such as generalization and construction are not readily modeled within this framework. But we can generate a degenerate macrostructure using only the selection operation. For each section we can find the most typical sentence in the section. For this purpose, we define *most typical* as the sentence with the highest average cosine to all the other sentences in the section. This will not always yield the best result because the ideal macroproposition may involve generalization or construction, but it will serve as a reasonable approximation.

Thus, some progress can be made toward a computational model of macrostructure generation. LSA allows us to generate an abstract vector representation of the macrostructure of a text (at least in those cases where the subsections of the text are clearly indicated, as in the example above). Furthermore, procedures can be devised to select for each section of a text the most typical sentence. However, that does not make a summary yet, and the operations for reducing the selected typical sentence to an essential phrase or fragment depend on more analytic procedures that go beyond LSA.

There are other, more practical uses of LSA's ability to represent the content of a text mathematically and compare it with other texts. For instance, we can express the summary written by a student as a vector and compare it with the vector of the to-be-summarized text. If the cosine between summary and text is high, the summary has much the same content as the original text. On the other hand, if the cosine is low, the summary does not reflect the content of the original text. A system, called *Summary Street*, that employs this method to help students write better summaries has been used with considerable success in some classrooms (E. Kintsch et al., 2000; Wade-Stein & Kintsch, 2003). For instance, students in sixth-grade classes were routinely asked to write summaries of chapters of their science textbooks. The teacher assigned a text to be summarized, say, on energy sources (coal, wind, petroleum, etc.) or Meso-American civilizations (Incan, Mayan, or Aztec). Each text is usually composed of four or five sections, and the teachers wanted the content of each section to be covered in the summary. Furthermore, the teachers required the summary to be of a certain length, say, between 150 and 200 words. The students write their summaries on an interface that is much like a standard word processor and send them to the LSA system for analysis via the Web. The feedback is received almost immediately and involves a number of steps.

Content feedback indicates whether all sections of the text have been covered in the summary. For this purpose, the cosine between the student's summary and each of the sections of a text are computed. If a cosine is below a certain threshold value, the student is told that this section is not adequately covered in the summary. The student then has the option to look at the appropriate section of the text on the computer screen and add some material about this section to the summary. If the threshold is exceeded for all sections, the student is told that he or she has now covered all parts of the text. Because the length of the summaries is restricted to avoid extensive copying from the source texts, students are told how long their summaries are so far and which of their sentences may be redundant or irrelevant.

Summary Street has been shown to be effective in helping students to write better summaries. When summary writing was compared with and without system feedback (Wade-Stein & Kintsch, 2003), the analysis showed that students were willing to work harder and longer when given feedback. Indeed, their time on task more than doubled. Summaries written with content feedback received higher grades from the teachers. This was the case for difficult summaries, for which grades more than doubled, whereas for texts that were easy to summarize anyway, the use of the system had no significant effect. Finally, a transfer effect was observed. Students who had written a summary in the previous week with the help of the system wrote better summaries a week later even when they no longer had access to the feedback the system provided. They had learned something about how summaries should be written.

Inferences and Situation Models[5]

Text comprehension always goes beyond the text. The mental representations that readers construct—their understanding of the text—depend as much on what readers bring to the text, such as their goals, interests, and prior experience, as on the text itself. Readers must make inferences to construct situation models. But not all inferences in comprehension are alike.

Classification of Inferences

A distinction should be made between problem-solving processes on the one hand, where there are premises from which some conclusion is drawn (not necessarily by the rules of logic)—which may be justly called inferences—and knowledge-retrieval processes on the other hand, where a gap in the text is bridged by some piece of preexisting knowledge that has been retrieved (W. Kintsch, 1998). Both inferences proper and knowledge retrieval may be either automatic (and usually unconscious) or controlled (and usually conscious and strategic). This classification results in the 2-by-2 table shown in Table 5.

Retrieval adds preexisting information to a text from long-term memory. Generation, in contrast, produces new information by deriving it from information in the text by some inference procedure. Thus, while the term *inference* is suitable for information-generation processes, it is a misnomer for retrieval processes.

TABLE 5
A Classification System for Inferences in Text Comprehension

	Retrieval A	Generation C
Automatic Processes	bridging inferences, associative elaborations	transitive inferences in a familiar domain
	B	D
Controlled Processes	search for bridging knowledge	logical inferences

Based on Kintsch (1998).

A prototypical example for cell A, the automatic retrieval process that enriches the information in a text, would be the activation of *with a hammer* by *John nailed down a board*, or *cars have doors* by *A car stopped. The door opened*. In both cases sufficient retrieval cues for the information retrieved exist in short-term memory. These cues are linked with pertinent information in long-term memory. Such knowledge use is automatic and rapid, and it places no demands on cognitive resources.

There are two theories that describe automatic knowledge retrieval. One is the long-term working memory theory of Ericsson and Kintsch (1995), which is described in more detail in the next section of this chapter. According to the long-term working memory theory, for well-practiced associations, retrieval cues in short-term memory are linked to contents in long-term memory, which thereby become directly available, thus expanding the capacity of working memory. An alternative model for this kind of knowledge retrieval is the resonance theory of Myers (Myers, O'Brien, Albrecht, & Mason, 1994). According to this model, cues in short-term memory produce a resonance in long-term memory, so that the resonating items become available for further processing in working memory. Thus, either via retrieval structures or resonance, relevant, strongly related items in long-term memory become potential parts of working memory, creating a long-term working memory that is much richer than the severely capacity-restricted short-term working memory. Indeed, it is only this long-term working memory that makes discourse comprehension (or, indeed, any other expert performance) possible. Smooth, efficient functioning would be impossible if we had no way of expanding working memory capacity beyond the rigid limits of short-term memory.

In cell B of Table 5 are cases where automatic retrieval is not possible. That is, the cues present in short-term memory do not retrieve relevant information that bridge whatever gap exists in the text. An extended search of memory is required to yield the needed information. A memory search is a strategic, controlled, resource-demanding process in which the cues available in short-term memory are used to retrieve other likely cues from long-term memory that in turn are capable of retrieving what is needed. Consider the following sentences:

(19) *Danny wanted a new bike. He worked as a waiter.*

Purely automatic, associative elaboration might not retrieve the causal chain from *want-bike* to *buy-bike* to *money* to *work*. However, a directed search for causal connections between the two sentences would easily generate these by-no-means-obscure links. In all probability, genre-specific strategies exist to guide such search processes. In a story, one would look for causal links. In a legal argument, one routinely looks for contradictions. In an algebraic word problem, algebraic formulas guide the search. The difficulty of such procedures, and the resource demands they make, vary widely.

Retrieval processes merely access information available in long-term memory, either automatically or by a resource-demanding search. Generation processes actually compute new information on the basis of the text and relevant background information in long-term memory. They, too, may be either automatic or controlled.

Some generation procedures are fully automatic (cell C of Table 5). For instance, given the sentence

(20) *Three turtles rested on a floating log, and a fish swam beneath them.*

the statement *The turtles are above the fish* is immediately available to a reader. Indeed, readers often are unable to distinguish whether they were explicitly told this information or not (e.g., Bransford, Barclay, & Franks, 1972). Note, however, that this is not merely a question of knowledge retrieval as in *doors are parts of cars*: The statement *the turtles are above the fish* is not something that already exists in long-term memory and is now retrieved, but it is generated during the comprehension process. The reason why it is so highly available in the reader's working memory is, presumably, that the fish-log-and-turtle scene is encoded as an image, and this mental image constitutes a highly effective retrieval structure that provides ready access to all its parts—not just the verbal expression used in its construction.

The information that allows the reader to infer that the turtles are above the fish is, presumably, in the form of a spatial image. It is given directly by the image that serves as the situation model representation of the sentence in question. Indeed, at this level of representation there is no difference between explicit and implicit statements. A difference only exists at the level of the textbase and surface representation, which, however, may not always be effective (as in the experiments of Bransford et al., 1972, in which subjects could not distinguish between explicit and implicit statements, given study and test sentences as in the example discussed here).

However, what happens in cell C of Table 5 should hardly be called an inference either. It is simply a case, in which due to the analog nature of the mental representation involved, more information is generated in forming a situation model than was explicit in the text. The term *inference* really should be reserved for cell D of Table 5. This is the domain of deductive reasoning. It is a domain that extends far beyond text comprehension, although deductive reasoning undoubtedly plays an important role in text comprehension, too. Explicit reasoning comes into play when comprehension proper breaks down. When the network does not integrate, and the gaps in the text cannot be bridged any other way, then reasoning is called for as the ultimate repair procedure.

Inferences (real inferences, as in cell D) require specific inference procedures. It is a matter of considerable controversy in psychology what these inference operations are—whether inference proceeds by rule (Rips, 1994) or mental model (Johnson-Laird, Byrne, & Schaeken, 1992). Inferences in domains where the basic representation is an action or perceptual representation, that is, analog rather than linguistic or abstract, probably involve operations on mental models. Inferences in truly symbolic, abstract domains may be by rule. Inferences in the linguistic domain, where the representation is at the narrative level, may be based on mental models but also could involve purely verbal inference rules.

Inference Generation During Discourse Comprehension

The literature on "inferences" in discourse comprehension is for the most part not concerned with cell D of Table 5. Indeed, it is heavily concentrated on cell A, the processes that are the least inference-like, according to the argument presented here. A major focus of the recent research has been on the question of to what extent inferences are made during normal comprehension. On the one hand, it is clear that if the readers of a story are asked to make inferences and are given sufficient time and incentive, there is almost no limit to what they will produce (Graesser, 1981). On the other hand, there is good evidence that much of the time, and in particular in many psychology experiments, readers are lazy and get away with a minimum of work (e.g., Foertsch & Gernsbacher, 1994). McKoon and Ratcliff (1992, 1995) have elaborated the latter position as the *minimalist hypothesis*, which holds that the only inferences readers normally make are bridging inferences required for the maintenance of local coherence, and knowledge elaboration where there are strong preexisting, multiple associations. Many text researchers (e.g., Graesser & Kreuz, 1993; Graesser, Singer, & Trabasso, 1994; Singer, Graesser, & Trabasso, 1994), however, feel that this minimalist position underestimates the amount of inference-making that occurs during normal reading and would at the least add inferences that are necessary for global coherence to the list (superordinate goal inferences, thematic inferences, and character emotional reactions). While this controversy has contributed a great deal to our understanding of the role of inferences in text comprehension, it also has shown that the question concerning which inferences are necessary for and are normally made during text comprehension has no simple answer. Text characteristics (much of the research is based on stories, mostly ministories), task demands, and individual differences among readers create a complex, though orderly, picture.

Trabasso and Suh (1993) have combined discourse analysis, talk-aloud procedures, and experimental measures, such as recognition priming, reading times, coherence ratings, and story recall, to show that their readers did make causal inferences in reading a story and that these inferences could be predicted by their analysis.

In an illuminating series of studies, O'Brien and his colleagues have shown that causal inferences in story understanding should best be regarded as a passive operation that makes available background and causal antecedents via a resonance-like mechanism (or what I would call a retrieval structure). Such a process contributes to the coherence of the text representation (Garrod, O'Brien, Morris, & Rayner, 1990) but is not predictive. Readers refrain from prediction unless there is absolutely no chance of being discomfirmed (O'Brien, Shank, Myers, & Rayner, 1988). Global automatic goal inferences occur only under limited conditions (Albrecht, O'Brien, Mason, & Myers, 1995), probably because such inferences are as risky as predictions: They are frequently discomfirmed as the later text reveals a different goal. When global goal inferences occur, resonance describes what happens better than the notion of inference. Through resonance, related parts of a text are connected because of preexisting retrieval structures. In contrast, the construction of a full mental model with rich causal connections appears rather as a nonautomatic, controlled process (O'Brien, 1995; Albrecht & O'Brien, 1995).

How much time and resources the reader has strongly determines the amount of inference-making that occurs. Magliano, Baggett, Johnson, and Graesser (1993), using a lexical decision task, found that causal antecedent inferences were not made when texts were presented rapidly at a 250 msec rate, but they were made when the presentation rate was 400 msec. Long, Golding, and Graesser (1992) found that superordinate goal inferences linking various episodes of a story (but not subordinated goal inferences) were made by readers when they were given a lot of time. However, with a rapid presentation rate, only good comprehenders made such inferences, while there was no evidence for goal inferences by poor comprehenders (Long & Golding, 1993).

Readers are much more likely to make antecedent causal inferences than consequent causal inferences (e.g., Magliano et al., 1993). For instance, readers of *The clouds gathered quickly, and it became ominously dark. The downpour only lasted 10 minutes* infer the causal antecedent *the clouds caused the rain*. But given *The clouds gathered quickly, and it became ominously dark*, they do not infer the consequent *the clouds caused rain*. This finding that antecedent, but not consequent causal, inferences are made in text comprehension is readily accounted for by the CI model. Suppose a text describes a situation that is a common cause of some event and then asserts that this event occurred, without mentioning an explicit causal connection between the antecedent and the event. Preexisting retrieval structures causally link the antecedent and the event in the reader's memory, and the causal link will be activated and is likely to become a permanent part of the reader's episodic text memory because it connects two highly activated nodes in the memory structure.

The situation is different for the consequent inferences. The same retrieval structures that made available the causal antecedent will make available the causal consequent, too. But at that point in the reading process, the consequent is a dangling node in the episodic text structure because it is connected to nothing else in the network but the antecedent. Therefore, the consequent will not receive much activation in the integration process and will be excluded from episodic memory. Thus, *The clouds gathered quickly, and it became ominously dark* might make available *the clouds caused rain*, but if nothing else in the text connects to *rain*, this node will become quickly deactivated in the network. When in a later processing cycle other information becomes available that could have linked with *rain*, that node is most likely lost from working memory. Hence, although the retrieval structures in the reader's long-term memory make available both antecedent and consequent information, only the former is likely to survive the integration process and become a stable component of the reader's text memory.

Time Course for Constructing Knowledge-Based Inferences

Of considerable interest is the time course of constructing knowledge-based inferences in text comprehension. We know that it takes about 300–350 msec for word meanings to become fixed in a discourse context. Inferences require more time. In Till, Mross, and Kintsch (1988), no evidence for topic inferences was obtained at a Stimulus Onset Asynchrony (SOA; the time interval between the presentation of the target word and the test word) of 500 msec, but topic inferences were clearly made at an SOA of 1,000 msec (there were no data points in between). In contrast, Magliano et al. (1993) found that antecedent causal inferences required an SOA of only 400 msec. Long, Oppy, and Seely (1994), in a study modeled after Till et al., have used SOAs of 200, 300, 400, 500, 750, and 1,000 msec. Associative effects are already fully apparent in their data at 300 msec. Topic effects develop gradually: They are already apparent at 500 msec but increase in strength up to 750 msec. Because different materials and conditions were used in all these studies, the differences in the results are not surprising. It seems that sentence-level inferences require from 400 to 750 msec, depending on experimental conditions. Thus, sentence meanings take roughly twice as long as word meanings to fixate.

The Construction of Situation Models

Much recent research has been concerned with the construction of situation models (e.g., Glenberg, Kruley, & Langston, 1994; Glenberg & Langston, 1992; Graesser & Zwaan, 1995; Mani & Johnson-Laird, 1982; Trabasso & Suh, 1993; Zwaan, Magliano, & Graesser, 1995). There is no single type of situation model and not a single process for the construction of such models. Situation models are a form of "inference" by definition, and Table 5 is as relevant for situation models as it is for any other "inference" in discourse comprehension. That is, situation models may vary widely in their character. In the simplest case, their construction is automatic. Relevant information is furnished by existing retrieval structures, as in the examples given for cell A in Table 5. Or it may be available simply as a consequence of a particular form of representation, such as imagery. Such situation model inferences do not add new propositions to the memory representation of the text but simply make available information in long-term memory via retrieval structures, or information that is implicit in the mental representation, such as an image (see Fincher-Kiefer, 1993, and Perfetti, 1993, for similar suggestions). On the other hand, situation models can be much more complex and result from extended, resource-demanding, controlled processes. All kinds of representations and constructions may be involved. The process may be shared by a social group or even by a whole culture and extend over prolonged periods of time. Text interpretation is not something that is confined to the laboratory.

Spatial and temporal information are usually important components of a situation model. Perrig and Kintsch (1985) had subjects read descriptions of the spatial layout of a small town. The same town was described in two ways, first by providing route descriptions (*after the church, turn right on Main Street to go to the courthouse*), and second by means of survey descriptions (*the courthouse is north of the church on Main Street*). Subjects were tested both for their ability to recall and recognize the text they had read and to make novel spatial inferences on the basis of that text. The results of their first experiment dramatically illustrated the textbase–situation model distinction: Subjects' recall was excellent and sentence recognition nearly perfect—but their ability to verify inferences was similar to results of random choices. In a second experiment, with a simpler town and more study time, subjects successfully constructed a spatial situation model. They performed well on recall and recognition as well as on inference tasks. Interestingly, the kind of situation model differed, depending on the text they had read: Route texts led to route models, and survey texts led to survey models. When the inference question was in the same form as the text a subject had read, performance was better than when the text was a route description and the question in the survey format or vice versa. The Perrig and Kintsch study shows that situation models are by no means automatic consequences of good textbases and that there may be different types of situation models. Which one is best depends on the reader's purpose.

A study by van der Meer, Beyer, Heinze, and Badel (2002) explored the construction of temporal situation models by presenting events in their chronological order (*fall down–get up*) or in reverse order (*fall down–slip*). Overall, chronologically related information was accessed faster compared with reversely ordered sentences, but processing time made a crucial difference. When there was not enough processing time, neither chronological nor reverse information was integrated into the situation model. When there was a great deal of time for elaboration, both were integrated. In the intermediate condition, however, chronologically ordered events were integrated into the situation model, whereas reverse, past-oriented events were not. Thus, what sort of inferences people make and how elaborate a situation model they construct depend crucially on the amount of processing. If there is time and they are motivated, people will construct rich situation models—but that is a controlled, effortful process, not the kind of automatic knowledge activation discussed earlier. We shall return to the educational implications of this finding in a later section.

The Role of Working Memory in Comprehension[6]

Within the standard working-memory framework, it is not possible to explain how memory is used in many cognitive tasks such as playing chess or text comprehension. The span of immediate memory is seven items plus-or-minus two. People can maintain no more than three or four discrete items in consciousness at the same time. Yet text comprehension requires that people juggle in their working memory large amounts of information: perceptual features, linguistic features, propositional structure, macrostructure, situation model, control structure, goals, lexical knowledge, frames, general knowledge, and episodic memory for prior text. Each of these components by itself would exceed short-term working memory, but each is clearly needed for text understanding, and people have no memory problems in understanding well-written, familiar texts.

Similarly, people can remember well-written, familiar text very well and effortlessly, just as they can remember what they did last night. In the laboratory, on the other hand, it takes people one hour to memorize 100 random words, and from a list of 30 words, college students recall about 12 or 14 after one reading. So why is memory so poor in the lab and so good in (some) everyday situations? Is our psychology of memory irrelevant to real-life memory and comprehension?

The theory of long-term working memory (LTWM) addresses this dilemma. It does so by specifying the conditions under which working memory capacity can become greatly expanded and by describing the mechanisms that are responsible for this expansion of working memory. The theory was first proposed by Ericsson and Kintsch (1995) and further elaborated by W. Kintsch (1998) and W. Kintsch, Patel, and Ericsson (1999).

LTWM is restricted to well-practiced tasks and familiar knowledge domains. With novel tasks and in unfamiliar domains, people must do with short-term working memory, whose capacity is severely restricted. Because the typical laboratory tasks—such as memorizing a list of paired associates—were unfamiliar to the subjects of memory experiments—and the materials used were relatively meaningless—word lists, or, in the extreme case, nonsense syllables—most laboratory studies of memory never involved more than short-term working memory; hence, the ubiquitous findings of severe capacity limitations. However, in some real-life situations in which people perform tasks at which they are highly skilled and well practiced, performance does not suffer from memory limitations. Skilled, expert performance provides many examples of such situations—playing chess or medical diagnosis, for instance. Of course, not everyone playing chess will have a memory advantage. Only the real expert shows exceptional memory in such tasks. Novice chess players can remember briefly presented chess positions no better than the capacity limitations of short-term memory allow them. Only master chess players who have devoted a decade or so to the study of chess will show truly superior memory in these situations. Indeed, part of becoming an expert in a skill consists in the development of superior memory in the expert domain. These memory skills are entirely domain specific, however. The chess master on all memory tasks outside his or her expertise performs no better than people normally do.

Thus, LTWM is an expert skill. There are, however, tasks at which most adults in U.S. society are experts. Text comprehension is an example. As long as the texts to be comprehended are simple, reasonably well written, and about familiar, everyday topics, we are all experts. The reading (or listening) skills involved are highly practiced over a lifetime. The subject matter of many texts often concerns everyday events and human actions and relationships—subjects in which our lifelong experience qualifies us as experts. A text on atomic physics needs a physicist to comprehend, but for a simple story or item in the newspaper, we all have the necessary expertise. Thus, we comprehend such texts readily, retrieve relevant knowledge or personal experiences automatically without special effort, and remember what we read, also without special effort. The LTWM mechanism is responsible for this achievement and explains why our memory is so good in familiar domains and so poor when we read something in an unfamiliar domain or try to acquire a new skill.

The LTWM theory claims that superior memory in expert domains is due to LTWM, whereas in nonexpert domains LTWM can be of no help. Thus, working memory has two components: (1) short-term working memory, which is available under all conditions but is severely limited in its capacity, and (2) LTWM, which is not capacity limited but available only in expert domains. Short-term working memory is what has been studied in most laboratory memory tasks. LTWM is conceived as a subset of long-term memory[7] that is directly retrievable via cues in short-term working memory. Any cue in short-term memory—alternatively we could talk about the contents of consciousness, or items in the focus of attention—that is linked by a stable memory structure to long-term memory nodes makes available these nodes in a single automatic and quick retrieval operation. The retrieval is fast and automatic in that it does not require mental resources (such as an intentional, conscious memory search does). Thus, the contents of short-term memory automatically create LTWM: a zone in long-term memory that is directly linked to these contents and immediately retrievable. The crucial restriction is that the items in short-term working memory and the items in long-term memory are linked by stable, fixed memory structures that permit direct retrieval. This is the case only in very well-practiced domains in which we are experts. Without these expert memory links, retrieval can be a protracted and resource-demanding process and is controlled rather than automatic.

Long-term memory is a relatively permanent system. Additions and modifications occur, as well as forgetting, but the system as a whole changes slowly. Short-term memory, the focus of attention or content of consciousness, on the other hand, changes from moment to moment. Because LTWM is generated dynamically by the cues that are present in short-term memory, LTWM mirrors the changes in short-term memory. A flashlight metaphor often has been used to describe short-term memory: a small beam that lights up three to five nodes in long-term memory. Imagine each of these nodes is linked to nodes in the unlit part of the long-term memory network. The linked nodes form LTWM. Working memory consists of the lit nodes plus the linked nodes in the dark part of long-term memory. The flashlight is able to jump immediately to any of these linked nodes, without external guidance.

The previous description represents the simplest case of LTWM. The links [stable associations or other memory structures such as schemata, frames, etc.] preexist in long-term memory. LTWM in this case involves no more than a set of cues in short-term memory plus the long-term memory nodes to which they are linked in long-term memory. But this is only part of the story because the ongoing cognitive process results in the generation of new nodes, which greatly enrich and complicate LTWM. These nodes are first generated in short-term working memory, but as the focus of attention shifts away, they fade from consciousness. Depending on the nature of these nodes, they may be more or less permanent or subject to forgetting.

Consider what happens in reading comprehension: Comprehension results in the formation of new nodes in memory (propositions derived from the text) that are linked in a complex pattern determined by the nature of the text and the comprehension strategies of the reader. The newly formed links in text comprehension are the result of the reader's comprehension strategies, as specified by the CI model. Some links are strong, some are weak, some nodes are tightly interconnected, and some are sparsely interconnected, depending on how the mental representation of a text has been built up. The structure that supports retrieval is not being formed for the purpose of memory retrieval. Rather, the ability to retrieve is incidental to comprehension: If one comprehends a text properly, a mental structure has been generated that supports memory retrieval via LTWM. What is required for LTWM, therefore, are appropriate comprehension strategies (e.g., as described in van Dijk & Kintsch, 1983) and the knowledge (linguistic knowledge, world knowledge, specific topic knowledge) and skills (language skills) necessary for the use of these strategies.

LTWM is not always incidental. There is a continuum between some processes where LTWM is incidental, as in text comprehension or chess, and other processes where it is intentional, as in the case of the retrieval structures used in mnemonic techniques. Thus, the memory artist studied by Chase and Ericsson (1981) employed a set of specific encoding strategies for digit strings for the sole purpose of memorizing them and used a body of knowledge (about running times) that was needed for the operation of these encoding strategies. Another example is the method of loci, where a complex schema is used over and over again, together with specialized imagery encoding strategies, for the sole purpose of memory retrieval. It is important to realize, however, that the deliberate retrieval structures involved in mnemonic techniques are but one type of structure that supports LTWM. Incidental structures that arise from text comprehension processes or planning moves in a chess game represent quite different cases that are ecologically more important.

Two types of links are involved in LTWM: (1) links among newly formed nodes as a text is being comprehended and (2) links between these newly formed nodes and other nodes in long-term memory. For new nodes and links in text comprehension, the process assumed here is the following: Certain features of a text elicit an appropriate processing strategy; the application of this strategy results in the creation of new memory nodes, links among them, and links among the new nodes and the body of long-term memory. The whole process is automatic. Thus, faced with a particular text, an expert speaker of the language automatically recognizes which comprehension strategies are appropriate, applies them, and generates a network of propositions linked to prior knowledge. A chess player looks at a board and applies appropriate planning strategies, creating a network of representations that enable later recall. A memory artist "sees" a random digit string as a meaningful running time and stores it at a particular place in his reusable retrieval structure. An expert physician recognizes a patient's signs and symptoms, which are stored as a pattern for subsequent decisions about disease, therapy, and management. The performance is quick and effortless in each case but limited to the specific domain in question. In each case, an episodic memory structure is created that supports LTWM. The nature of these strategies, and the resulting structures, is the object of study in psycholinguistics, the psychology of chess, the psychology of clinical reasoning, or mnemonics, respectively, and differs widely between these domains. Although we know something about these strategies and their use, much remains to be learned.

When the text is familiar and the reader is experienced, retrieval structures are thus generated automatically as incidental results of the comprehension process. Student readers, on the other hand, cannot rely on LTWM because they lack the automatic strategies and the background knowledge that makes it possible. The goal of instruction must be to get the student engaged in comprehension processes that are equivalent to the comprehension strategies experts employ that result in retrieval structures and the buildup of LTWM. However, what is automatic and effortless for the expert reader is intentional and effortful for the novice. The theory of LTWM implies that student readers cannot read in the same way as experienced readers, but instead must be taught explicitly to engage in the right kind of comprehension strategies. They have to be induced to be active readers, even though that may be hard work for them, because otherwise their comprehension will be superficial and their memory poor, and their ability to integrate information from a text with their background knowledge remains severely limited. In the next section, the implications of this finding will be explored for learning from text.

Learning From Text[8]

There are important psychological differences between learning from a text and remembering the text. Text memory, that is, the ability to reproduce the text either verbatim, in paraphrase, or by summarizing it, may be achieved on the basis of only superficial understanding. In the extreme case, one can learn to recite a text by rote without understanding it at all. Learning from text, on the other hand, requires deeper understanding. Learning from text is the ability to use the information acquired from the text productively in novel environments. This requires that the text information be integrated with the reader's prior knowledge and become a part of it, so that it can support comprehension and problem solving in new situations. Mere text memory, on the other hand, may remain inert knowledge—reproducible given the right retrieval cues, but not an active component of the reader's knowledge base. Text memory is based on the textbase; learning from text requires the construction of a situation model.

A well-known study by Bransford and Franks (1971) can be characterized as all understanding and no memory. In this study, the texts consisted of four simple sentences such as *The ants were in the kitchen. The ants ate the jelly.* and so on. These ideas also could be expressed in more complex sentences such as *The ants in the kitchen ate the jelly.* Subjects were given several such texts, in either the form of 4 one-idea sentences, 2 two-idea sentences, 1 three-idea sentence and

1 one-idea sentence, or other such combinations. Later, they were given a recognition test consisting of some sentences they had actually read and others they had not seen before. The results of the study were clear: Subjects remembered very well the stories they had read (e.g., they remembered the ants and the jelly and whatever else there was to that text) but did not know which particular sentences they had read. They remembered the meaning of each minitext—a scene or an image, perhaps—but not the way it had been presented verbally. The memory for the actual text they had read was wiped out by heavy interference (the subjects read many sentences, all very similar), but they had no trouble keeping in mind the few simple and distinct situation models they had formed for each of the several texts they had read.

At the other extreme—all memory and little understanding—is a study by Moravcsik and Kintsch (1993). Three factors were varied in their experiment: (1) domain knowledge (high or low), (2) the way the text was written (good, well-organized writing vs. poor, disorganized writing), and (3) reading ability (high or low scores on the comprehension subtest of the Nelson–Denny Reading Test). Domain knowledge was manipulated by using titles that allowed subjects to use their knowledge and to disambiguate the otherwise obscure text.

All three factors—knowledge, writing quality, and reading ability—significantly influenced the amount of reproductive recall. These effects were additive. There was no indication of an interaction and, hence, of compensation (see also Voss & Silfies, 1996). However, there was an interesting difference in the kind of mental representation subjects constructed in the different experimental conditions. Even though high reading skill and good writing enabled low-knowledge readers to form adequate textbases that were capable of supporting reproductive recall, these readers could not form correct situation models to support their elaborative recall.[9] Their elaborations tended to be wrong and fanciful. Only high-knowledge readers were capable of good elaborations. Inadequate situation models did not keep the subjects from elaborating, but their elaborations and inferences were erroneous, whereas the elaborations of readers who could use their knowledge to construct an adequate situation model were appropriate ones.

It is interesting to note the effects of writing quality in this study. All passages were written in two different versions, preserving their content but varying their style. In one version, the language was as helpful as my colleagues and I could make it in signaling discourse importance to the reader. The other version was as unhelpful as my colleagues and I could make it while still writing an English text. Writing quality had a major effect on reproductive recall, facilitating reproduction about as much as domain knowledge did. But it did not help understanding: Whether a text was well written did not have a statistically significant effect on the proportion of erroneous elaborations. Thus, while good writing can help the reader to construct a better textbase, sufficient for recall, it does not by itself guarantee the deeper understanding that is a prerequisite for learning.

Just how low-ability students go about using their domain knowledge to achieve good comprehension results was investigated by Yekovich, Walker, Ogle, and Thompson (1990). One group of their subjects had high knowledge of football, and one had low knowledge. With texts that had nothing to do with football, the two groups performed equivalently. But with a football text, the high-knowledge subjects outperformed the low-knowledge subjects. The questions on which the high-knowledge subjects showed the greatest advantage over low-knowledge subjects were inference questions and integrative summary statements. There was less of a difference on memory and detail questions. This is, of course, exactly what the theory of comprehension would lead one to expect. Even readers with little domain knowledge can understand information that is given explicitly in the text (although they might not remember it because their retrieval structures might not be rich enough). However, inferences and thematic integration that build retrieval structures require knowledge.

The Measurement of Learning

In order to be useful in research, the theoretical distinction between learning from a text and memory for a text requires empirical methods to assess learning separately from memory. However, because learning and memory cannot be separated cleanly even in the theory (textbase and situation model are not two separate structures but are the text-derived and knowledge-derived components of a single structure), measurement procedures are not precisely separable into textbase and situation model measures either. Instead, empirical measures reflect one or the other aspect of the structure to a stronger degree. Thus, one can ask questions that demand a specific detail from the text or that require the integration of textual information and prior knowledge in order to solve a new problem. Even recall reflects both aspects: the textbase, to the extent that the recall is reproductive, and the situation model, to the extent that it is reconstructive. Usually, recall is a mixture of the two, but in some cases it is primarily one or the other.

Text memory is measured through free recall, cued recall, summarization, various types of recognition tests, and text-based questions. Different methods are needed, however, for the measurement of learning. Psychology shares the need for such measures with artificial intelligence (AI), in so far as AI is interested in the construction of knowledge-rich expert systems (e.g., Olson & Biolsi, 1991), and with education, where the assessment of learning is of obvious importance. Education and AI have relied for the most part on direct methods for knowledge assessment, that is, various forms of question asking. That is still by far the most widely used method in psychology too, although more indirect scaling methods also have been developed for purposes of psychological research.

Asking questions is a method for the assessment of knowledge that is fraught with problems. Developers of expert systems rely on this method almost exclusively, but it is difficult to determine the correctness or completeness of the answers that are being elicited. Educationally, the problem is that asking questions is artificial and sometimes yields invalid results. It is an unnatural act when a teacher asks a student for something she knows better than the student. Furthermore, the answers that the students give may indicate much else other than real learning. Students may acquire specific strategies that allow them to generate acceptable answers without deeper understanding. Or questions may be answered correctly or wrongly for various accidental reasons that have nothing to do with the students' understanding. These problems are widely appreciated but not easily avoided.

Instead of direct questions, scaling methods often are employed to assess knowledge indirectly. Scaling methods require a set of keywords or phrases that are characteristic for a certain knowledge domain. (One can ask experts for such words or use more objective methods such as frequency counts of technical terms in relevant scientific publications.) The knowledge of a subject is inferred from the way the subject organizes these keywords. If the structure generated by the subject resembles the structure generated by domain experts, we infer that the subject's knowledge organization is similar to that of the experts. To the extent that the subject structures the set of keywords in ways that differ from the experts, a lack of correct domain knowledge is revealed.

The basic technique for finding out how a subject organizes a set of keywords is to ask the subject for relatedness judgments between all pairs of keywords in the set. A similarity matrix between all keywords is thus obtained, showing the rated closeness between all word pairs. Such a matrix can then be used as the basis for multidimensional scaling (e.g., Bisanz, LaPorte, Vesonder, & Voss, 1978; Henley, 1969). A low-dimensional space is generated in which the keywords are embedded. One can then ask whether the space is the same for the students as for the experts and whether the location of the keywords in this space is similar for students and experts. This method has been used successfully a number of times. For instance, the semantic field of "animal names" has been scaled in this way, yielding a space with the two dimensions of size and ferocity, which account for 59% of the variance of the paired-comparison judgments (Henley, 1969).

However, these scaling methods are of limited usefulness. The pairwise comparison method is laborious for the subject and rapidly becomes impossible to use as the number of keywords increases. Furthermore, multidimensional scaling methods work with group data, but we often need to work with data from individual subjects. Most important, however, it has become apparent that very few knowledge domains (other than animal names) are so regular and simple that they could be described by a space of a few namable dimensions.

A method for indirect knowledge assessment was developed by Ferstl and Kintsch (1999). It was applied to the problem of measuring the amount of learning that occurred from reading a text. If the text had an effect on the reader's memory, knowledge, or both, it should change the way the reader organizes a knowledge domain, and the change should be in the direction of the text organization. A reader's knowledge about a particular domain of interest is first assessed, either by a pairwise comparison of appropriate keywords or by collecting cued associations to these keywords. The reader's knowledge organization can then be inferred from these data either with the hierarchical clustering analysis of Johnson (1967) or the Pathfinder analysis of Schvaneveldt (1990). The reader then studies the to-be-learned text. Afterward, the reader's knowledge organization is reassessed to see whether it changed in accordance with the text organization.

Using Coherent Text to Improve Learning

Learning from text requires that the learner construct a coherent mental representation of the text and that this representation be anchored in the learner's background knowledge. Thus, one reason why students might fail to learn something from reading a text could be that they are unable to form a coherent textbase linked to their prior knowledge. It is easy to see why this might be the case for low-knowledge readers. Not all links either within a text or between the text and the readers' knowledge are always spelled out in a text; they are often left for the reader to fill in, for example, as *bridging inferences*. This is fine and, as we shall see in the next section of this chapter, can be quite advantageous, but it often creates problems for low-knowledge readers. If readers simply do not have the necessary background knowledge to fill in the gaps in the text that an author has left, they will be unable to form a coherent representation of the text or to link it with whatever little they do know. Consider some trivial examples:

(20) *The heart is connected to the arteries. The blood in the aorta is bright red.*

For a reader who does not know what the relationship between *arteries* and *aorta* is, there will be a coherence problem. Or consider the following example:

(21) *To stop the North Vietnamese aggressors, the Pentagon decided to bomb Hanoi.*

This sentence may present all kinds of problems to a low-knowledge reader: Namely, what is the *Pentagon* and how does *Hanoi* get into this sentence? A little rewriting of these problem texts that makes the relations between items in the text, or between general knowledge and the text, fully explicit can avoid these problems. Thus, we might define *aorta* for a low-knowledge reader by rewriting (20) as shown at the top of the next page.

(20a) *The heart is connected to the arteries. The blood in the aorta, the artery that carries blood from the heart to the body, is bright red.*

And we can help a low-knowledge reader with (21) by inserting explicit links between the unknown terms and what a reader might be expected to know:

(21a) *To stop the North Vietnamese aggressors, the U.S. Defense Department in the Pentagon decided to bomb Hanoi, the capital of North Vietnam.*

So far we have only considered local coherence problems, but the global coherence of a text often also can be made more explicit. The macrostructure of a text is not always explicitly signaled in the text but may be left for the reader to deduce. This is fine for knowledgeable readers but can be a major source of confusion when the requisite background knowledge is lacking. Thus, most of us need no help to understand the structure of a four-paragraph text, with each paragraph describing the anatomical details of one of the chambers of the heart. However, for a reader who does not know that the heart has four chambers, a title such as *The four chambers of the heart*, plus appropriate subtitles or clearly marked topic sentences for each paragraph, can be of great help.

It is indeed the case that such relatively minor revisions of texts that ensure coherence at both the local and global level facilitate text memory as well as learning for readers who lack background knowledge. A representative study is one by Britton and Gulgoz (1991). The authors used a history text as their learning material that described the U.S. air war in Vietnam. The text was written at the time of the war and presumed considerable prior knowledge on the part of the reader—knowledge that at that time was probably readily available among the population to which the text was addressed. Many years after the war, the people who participated in Britton and Gulgoz's study had very little specific information about the war, and hence found the text hard going. Although they were able to recall quite a bit from the text, and answered more than half of the fact questions correctly, their performance was poor on the inference questions. In fact, they really did not understand the text at all. This conclusion follows from an analysis of the conceptual understanding of the situation as it was assessed by a keyword comparison task.

Britton and Gulgoz selected 12 keywords that were of crucial importance for the understanding of the text and had students give pairwise relatedness judgments for these keywords. They also had the original author of the text plus several other experts on the Vietnam War provide relatedness judgments for the same set of keywords. The author and experts agreed quite well among themselves (average intercorrelation $r = .80$). But the students who had obtained their information from reading the text did not agree with the author or the experts at all (average $r = .08$). Their understanding of the air war in Vietnam on the basis of reading this text was quite different from the one the author had intended (and which the experts achieved from reading the same text). Thus, even though the students recalled a good part of the text and answered questions about it reasonably well, they really did not understand what they were saying. Britton and Gulgoz also report an analysis of the keyword judgment data, which make clear some of the fundamental misconceptions of these readers. For instance, whereas the text emphasized the failure of Operation Rolling Thunder, in the students' judgments *Rolling Thunder* was linked to *success*, instead. Britton likens this result to reading the Bible and concluding the devil was the good guy.

This dismal performance could be improved significantly by some rather simple revisions of the original text that make the text understandable to readers without adequate background knowledge. Britton and Gulgoz located all coherence gaps in the original text that required readers to make a bridging inference. Then they inserted into the text a sentence or phrase making this inference explicit. For instance, if in one sentence *North Vietnam* is mentioned and the next sentence begins with *In response to the American threats, Hanoi decided*, Britton and Gulgoz might have made this sentence pair coherent by revising the second sentence to *The North-Vietnamese government in Hanoi decided....* These revisions were highly effective. Recall increased significantly, as did the performance on inference questions. Readers now understood the text more or less correctly. Their relatedness judgments after reading the revised text correlated reasonably well with those of the experts ($r = .52$), and if one looks at the structure of their judgments as revealed by a Pathfinder analysis, no glaring misunderstandings are apparent, as was the case for the readers of the original version of the text. Hence, making this text locally coherent by filling in the gaps that required bridging inferences yielded a text that readers could understand, even though their background knowledge was lacking.

Several other studies confirm and extend these results (e.g., Beck, McKeown, Sinatra, & Loxterman, 1991; McKeown, Beck, Sinatra, & Loxterman, 1992; McNamara, E. Kintsch, Songer, & W. Kintsch, 1996, Experiment 1). Revising a text for coherence is clearly an effective technique to further understanding and learning. On their own, students without adequate background knowledge cannot fill in gaps in the text that readers with greater familiarity with the domain bridge effortlessly and, in fact, unconsciously. For such readers, providing explicit bridging material in the text, at both the local and global levels, is a prerequisite for understanding and learning.

If this is so, why do authors ever leave gaps in their texts? Why do we not write fully coherent, explicit texts all the time? The answer is that to write such a text is an elusive goal; a writer must always rely on the reader's knowledge to some degree. There is no text comprehension that does not require the reader to apply knowledge: lexical knowledge, syntactic and semantic knowledge, domain knowledge, personal experience, and so on. The printed words on the page, or the sound waves in the air, are but one source of constraints that must be satisfied. The reader's knowledge provides the other. Ideally, a text should contain the new information a reader needs to know, plus just enough old information to allow the reader to link the new information with what is already known. Texts that contain too much that the reader already knows are boring to read and confusing (e.g., legal and insurance documents that leave nothing to be taken for granted). Hence, too much coherence and explication may not necessarily be a good thing.

Improving Learning by Stimulating Active Processing

The contention that if readers possess adequate knowledge, a fully explicit text is not optimal for them, was explored by McNamara et al. (1996). Such readers will remember more and learn better from texts that require them to assume a more active role in comprehension. Specifically, my colleagues and I hypothesized that the results obtained by Britton and Gulgoz, and others described previously, pertain to low-knowledge readers. Readers with good domain knowledge might react in very different ways.

In the McNamara et al. experiment, college students studied a middle school–level encyclopedia article on heart disease. Two versions of the text were used. In one case, the text was maximally coherent and explicit at both the local and global level. For this purpose, potentially ambiguous pronouns were replaced with full noun phrases; elaborations that linked unfamiliar concepts to familiar ones were added, as were missing sentence connectives; and a concept was always referred to in the same way (rather than by a synonym or paraphrase). In addition, titles and subtitles were used to indicate the macrostructure of the text, and explicit macropropositions marked the role of each paragraph in the text. The resulting version was the high-coherence text. The low-coherence version was constructed by deleting all these signals but not otherwise changing the content of the text.

The prior knowledge of the subjects was assessed via a knowledge assessment test consisting of a series of questions on the basic anatomy and functioning of the heart and a recognition test in which subjects were asked to match parts of the heart to a cross-sectional diagram of the heart and its major blood vessels.

After taking this test, subjects read either the high- or the low-coherence text twice and then responded to a series of posttests. First they were asked to recall the text in their own words. Then subjects were given four types of questions: (1) text-based questions, (2) elaboration questions that required relating text information to the reader's background knowledge, (3) bridging inference questions that required connecting two or more separate text segments, and (4) problem-solving questions that required applying text information in a novel situation.

The results of this study reveal a strong interaction between the level of prior knowledge of the students, the coherence of the text, and the method of testing. High-knowledge students always perform better than low-knowledge students. But when tests are used that assess primarily text memory, the high-coherence text is better for all types of students. However, when tests are used that depend on the construction of a good situation model, the low-coherence text actually yields better results for high-knowledge subjects than the high-coherence text. For low-knowledge subjects, on the other hand, the usual superiority of the high-coherence version over the low-coherence version is observed.

It appears that the high-coherence text that spelled out everything these readers already knew quite well induced an illusory feeling of knowing that prevented them from processing the text deeply. They were satisfied with a superficial understanding, which was good enough for recall and answering text-based questions, but they failed to construct an adequate situation model combining their prior knowledge with the information from the text. Hence they learned relatively little from the text that should have been the easiest and most effective.

These results were replicated by McNamara and Kintsch (1996) with the Vietnam War text of Britton and Gulgoz (1991) and a different knowledge manipulation: Half the subjects received a brief lesson about the Vietnam War before studying the text. For untrained, low-knowledge subjects, the original low-coherence text was quite ineffective, as was the revised, high-coherence text for the trained high-knowledge subjects. However, if readers who knew very little about Vietnam were given the coherent, well-written text, their knowledge organization was influenced by the text they had read. Similarly, the better-informed readers clearly learned more from the challenging text, as reflected by the way they sorted the keywords.

These results are paradoxical. Should we start writing incoherent texts and give disorganized lectures so that our better students will benefit from them? The answer to this question seems to be a qualified "yes." Making things too easy for a student may be a significant impediment to learning. However, just messing up a lecture is not a solution. Instead, we need to challenge the student to engage in active, deep processing of a text. This can be done, as we have shown here, by placing impediments in the path of comprehension, but impediments of the right kind and in the right amount. They must be impediments we have reason to think the student can overcome, if he or she tries hard enough, and the activity of overcoming these impediments must be learning relevant. We have shown previously that giving students a text with coherence gaps to study from, for which they do not have adequate background knowledge, is self-defeating. Such students need all the help they can get, and we need to organize and explicate the text for them as well as we can. But, as the literature on generation effects in memory also demonstrates (e.g., McNamara & Healy, 1995), students who are able to perform a task unaided should be encouraged to do so. They will remember better and learn better to the extent that the activity they are engaging in is task relevant. This was certainly the case in the experiments discussed previously in which the incoherent texts forced the high-knowledge students to establish local coherence relations on the basis of their own knowledge, to figure out the macrostructure of the text on their own, and to elaborate the textual material with what they already knew.

The fact that not just any self-generated activity is helpful for text memory has been shown in a series of studies discussed in McDaniel, Blischak, and Einstein (1995). Story recall can be improved by omitting occasional letters from words so that the reader must fill in the missing information from the context. Filling in missing letters forces readers to focus on the details of the stories that readers often disregard in favor of the story line. On the other hand, a different orienting task, such as reordering sentences, has no effect on story recall because readers of a story pay sufficient information to the relational information between sentences, even without the reordering task. These relations are reversed for descriptive texts. The reordering task helps recall by focusing the reader's attention on otherwise neglected order information, but the missing letters task has no effect because readers of essays are quite careful about the details anyway. Thus, simply placing obstacles in the reader's path that force them to expend extra effort does not benefit their learning. Instead, positive effects can be expected only if the extra processing they engage in is task appropriate.

If the effort is task appropriate, however, engaging the reader in active processing can be quite helpful. We have seen this for learning from text in the experiments above. This is also the case for sentence memory. There is a curvilinear relation between the strength of a causal connection between two statements and the memory strength of their connection measured by cued recall, indicating that neither too weak nor too strong links are elaborated as successfully as intermediate links (e.g., Myers, Shinjo, & Duffy, 1987; van den Broek, 1990). Analogous results have been reported by Battig (1979) for paired-associate learning, who reported better retention when the learning was made more difficult through intratask interference. There is also a literature on skill acquisition (Schmidt & Bjork, 1992) that shows that making the learning process too smooth is counterproductive. Learners acquiring a new skill must have the opportunity to face difficulties and learn to repair mistakes.

Mannes and Kintsch (1987) have reported a study that is similar in spirit to the experiments discussed here. In their study, readers were given an advance organizer that either fit perfectly with the target text and, hence, made it easy to read, or that structurally mismatched the target text, so that the readers had to engage in some cognitive effort to relate the advance information and the target text. The target text was a rather long article from a popular science publication about the industrial use of microbes. Because the students knew very little about microbes beforehand, an advance organizer was prepared that told them everything they needed to know about microbes. Two forms of this advance organizer were employed that differed in the order in which the material was presented but, as much as possible, preserved the content of the material. In one case, the general information about microbes occurred in exactly the same order as it was presented in the target text. We called this the congruent advance organizer. In the other version, the material was presented as in the encyclopedia article that we had used as our source. This organization was incongruent with the text.

Participants studied either form of the advanced organizer and took a short test on it, and then they read the target text, which required a certain amount of knowledge about the properties of microbes. After reading the text, they were tested on what they remembered from the text and what they had learned. A dissociation was observed between measures of text memory (correct verifications of old sentences) and measures of learning (correct verifications of inference sentences). When the advance organizer and the content of the text were structurally congruent, subjects readily understood the text and remembered it better than when the advance information and the text were organized differently. The formation of a textbase was presumably facilitated because the macrostructure they had formed for the advanced organizer provided a fitting schema for the text itself. When this was not the case, the textbase was not as perfect, and behavioral measures depending primarily on the textbase were lower. However, the subjects' situation model was enhanced because, in understanding the target text, they had to retrieve and integrate information from the advance organizer that had presented the information in a different context. Therefore, a richer, more interrelated network was constructed, which later helped the readers with inference questions and problem solving.

Thus, it has been demonstrated repeatedly that increasing the difficulty of the learning phase can have beneficial effects. Moreover, these studies are not limited to learning from text but involve other kinds of learning situations as well. Task difficulty can stimulate active processing, with the result that a more elaborate, better-integrated situation model will be constructed. However, this statement must be carefully qualified. The student must possess the necessary skills and knowledge to successfully engage in the required activity, and the activity must be task relevant.

Matching Readers and Texts

Readers without adequate background knowledge have trouble comprehending difficult texts and learn very little from them. On the other hand, the research just discussed shows that readers should be challenged and that making things too easy for students elicits a passive attitude that is not conducive to deep understanding and knowledge building. The problem is to find texts of the right level of difficulty for each student in a classroom where there is typically a broad range of preparation and prior knowledge. Experienced teachers and librarians can do that very well, if they have the time for each individual student. A procedure to do so automatically, proposed by Wolfe et al. (1998), has shown some promise in the laboratory but has not been classroom tested so far.

The procedure employed the ability of LSA to measure automatically how similar two texts are. Wolfe et al. asked students of varying background knowledge to write a brief essay about a medical topic. These essays revealed how much each student knew about the topic: Low-knowledge college students wrote quite different essays than students who happened to know a lot about the topic, and medical students who had studied this topic used a different language yet. Wolfe et al. had four texts, all about the same topic but varying greatly in sophistication, from a popularization for high school students, to college-level texts, to a medical school text. The participants in this study were randomly assigned to one of these instructional texts. How similar the text was to the essay the student wrote was measured by LSA: The cosine between the student's prior essay and the study text he or she was assigned provided a measure of the semantic distance between essay and text. After studying the text they were assigned to, the students were tested (by means of a questionnaire as well as another essay) for how much they had learned. When the cosine between a student's prior essay and the text was low—that is, essay and text were very different in content—there was little learning. When the cosine between essay and test was high—that is, essay and text were similar—there also was very little learning. However, for intermediate cosines (around .5), students learned quite well (30 to 40% improvement in their scores from pretest to posttest).

What Wolfe et al. found was an automatic and easy way to determine the "zone of learning" for individual students. The essay the student writes reveals the student's knowledge level. If the instructional text is too hard for that knowledge level, the student will learn very little from that text; if the text is too easy, there is not much for the student to learn. But if the student is given the right text for his or her level of preparation, learning can be successful. This is an intuitive result, but what makes the Wolfe et al. study interesting is the way they achieved it: Letting students express in their own words what they know about a topic and using LSA to index their knowledge quantitatively may turn out to be a useful practical method for various purposes. Not only can this method be used to match students and texts, as was done in their study, but it also is a useful tool for self-assessment. After reading an instructional text, students could write about what they learned and receive LSA-based feedback about the adequacy of their understanding, such as what content areas they had missed or irrelevancies they had introduced.

Word Problems: From Text to Action

Word problems—in grade school arithmetic, high school algebra, and college physics classes—challenge students because they require not only the correct formal operations but also a correct understanding of the problem. It is useful to view word-problem solving from the standpoint of text comprehension. The student must be able to form the correct situation model before formal operations can be brought to bear on the problem. A second grader may know that $8 - 5 = 3$, but not whether to add or subtract when given

(22) *Joe has 8 red marbles.*
Tom has 5 blue marbles.
How many marbles does Joe have more than Tom?

What makes word problems special is that the usual kind of situation model is insufficient. It is not enough for the student to understand that Tom and Joe have marbles, the marbles are blue and red, and one boy has more marbles than the other; in order to do the arithmetic correctly, a student needs to extract from the text very specific information. The arithmetic set schema specifies what that information is. The text must be used to fill in the empty slots in that schema, resulting in the construction of a problem model. Thus, it takes more than just an informal situation model to determine the appropriate arithmetic operation to solve the problem. A formal arithmetic problem model must be constructed. Normal text comprehension strategies do not support the construction of a formal problem model; special training with arithmetic word problems is required, which is why word problems are so difficult.

Many problems that require formal methods for their solution require this distinction between situation model and problem model. Consider, for instance, a scientist who wants to use a graphics program to design a graph for a manuscript he is writing. His situation model may be a mental image of the graph he wants to construct, specifying its principal features; he could draw such a graph by hand. To use the graphics program, however, the desired image has to be transformed into a formal problem model that takes into account the constraints and requirements of the particular graphics program he is using. In doing so, some details have to be specified that were not considered explicitly in the scientist's situation model (such as font size, precise arrangement of the graph axes, and so on). Only when the scientist's ideas are translated into the format required by the graphics program can action result: In this case, the program designs the desired graph. The second grader is in much the same situation: She must learn just what information in the text is needed for the arithmetic problem model, and in what form it is needed.

Word Arithmetic Problems

W. Kintsch and Greeno (1985; see also W. Kintsch, 1998, chap. 10.1) have formulated a model that simulates the construction of an arithmetic problem model for commonly used word problems in grades 1 to 4. It is based on the set schema.

which specifies for a collection of objects their type (e.g., *marbles*), their quantity (e.g., *8*), their specification (*Joe* is the owner), and the relation of this set to other sets (e.g., in [22] the set of Joe's marbles is to be compared with the set of Tom's marbles).

Consider how this simulation would solve a simple word problem such as (23), which is called a *transfer problem*.

(23) *Joe had 3 marbles.*

Then Tom gave him some marbles.

Now Joe has 8 marbles.

How many marbles did Tom give him?

The first sentence provides a cue for making a set S_1:

S_1:	Objects:	*marbles*
	Quantity:	*3*
	Specification:	*owner Joe*
	Role:	*unknown*

The second sentence cues a second set schema, S_2, and allows the model to determine the relationship between these sets because it knows that transfer problems (*give* is a cue that this problem might be of that type) have a *start-set*, *transfer-set*, and *result-set*.

S_1: Objects: *marbles*	S_2: Objects: *marbles*
Quantity: *3*	Quantity: *unknown*
Specification: *owner Joe*	Specification: *owner Tom*
Role: *start-set*	Role: *transfer-set*

The third sentence completes the problem model.

S_1: Objects: *marbles*	S_2: Objects: *marbles*	S_3: Objects: *marbles*
Quantity: *3*	Quantity: *unknown*	Quantity: *8*
Specification: *owner Joe*	Specification: *owner Tom*	Specification: *owner Joe*
Role: *start-set*	Role: *transfer-set*	Role: *result-set*

With the problem model completed, the model acts: It uses a procedure called Add-On (which is what first and second graders use in this case) to calculate the answer: *Tom gave Joe 5 marbles.*

What good is such a simulation? First, it specifies exactly what information the student must extract from the word problem. That does not mean that the student should be taught to use a set schema—just taught what the needed information is.

Second, the simulation helps us to understand why a child makes certain errors (and hence offers an opportunity for corrective instruction). Consider the following word problem:

(24) *Mark and Sally have 7 trucks altogether.*

Mark has 2 trucks.

How many trucks does Sally have?

The usual answer that the children give (45%, in a study by DeCorte, Verschaffel, & DeWinn, 1985) is 7. The problem is that the children do not know the meaning of *altogether* and parse it as *each*. The number of trucks Mark owns is updated by the second sentence, but they think they already know the answer about how many trucks Sally has when they read the third sentence. When these children are asked to recall problem (24), they recall it the way they understood it as follows, not the way it was actually presented:

(25) *Sally has 7 trucks.*

Mark has 2 trucks.

How many trucks do Mark and Sally have?

What these children are doing is turning a difficult problem that they are unable to solve as stated into an easy problem that they know how to solve—a strategy not only second graders employ.

A third example of how simulations, such as the one by W. Kintsch and Greeno (1985), illuminate the difficulties children have with word problems concerns the role of keywords in problem solving. Consider the following two problems:

(26a) *Tom is 175 cm tall.*	(26b) *Tom is 175 cm tall.*
Jeff is 12 cm shorter than Tom.	*He is 12 cm taller than Jeff.*
How tall is Jeff?	*How tall is Jeff?*

Problem (26a) is much easier than (26b) because the keyword *shorter* suggests subtractions, whereas *taller* suggests addition. Therefore, the second sentence of (26a) sets up a bias in favor of subtraction, which is helpful, while the second sentence of (26b) sets up a bias in favor of addition, which interferes with the correct response to the question posed in the last sentence.

Algebraic Schemas

Things get more interesting and more complicated when we consider algebraic word problems, rather than arithmetic word problems, because the importance of situation models in algebraic word-problem solving is at least as great as in arithmetic word-problem solving. In a pioneering study, using extensive protocol analyses, Hall, Kibler, Wenger, and Truxaw (1989) have shown that competent college students reason within the situational context of a story problem to identify the quantitative constraints required for a solution. They use the text to build, elaborate, and verify a situation model from which they derive their solution. A variety of reasoning strategies are used by students to develop these situation models, which are by no means restricted to the algebraically relevant aspects of a problem. Integrating the dual representations of a problem at the situational and mathematical levels appears to be the central aspect of competence according to Hall et al., as indeed it was for children working on arithmetic word problems. However, while there are only a few arithmetic schemas that the theorist must consider, over 1,000 problem types are found in algebra texts (Mayer, 1981).

This complexity, however, is primarily linguistic rather than algebraic. A relatively small number of algebraic schemas can do all the work that is needed for the construction of the necessary formal problem models. The difficulty we encounter is in the construction of the problem model. A broad range of linguistic and general world knowledge is needed for that purpose. The 1,000 algebraic problem types are based on only a few algebraic schemas, however. Four schema types are needed to deal with the largest class of algebra problems—rate problems. Another class of problems involves physics, geometry, and schemas from other domains. This is essentially an open class. Examples are Newton's second law—the sum of all forces for a system in equilibrium must be 0; Ohm's law; the Pythagorean theorem; the formulas for the area and circumference of a circle and other geometric shapes; and so on. Finally, there is a third class of word algebra problems—number problems—which are not schema based but must be constructed from the text directly without the use of a schema.

The four rate schemas are all of the same form:

$$\text{Unit}_1 = \text{rate-of-Unit}_1\text{-per-Unit}_2 \times \text{Unit}_2$$

The units may be amount per time (a boat travels 4 km per hour), cost per unit (a pound of almonds costs $5.99), portion to total cost (7% of the cost of a car), and amount to amount (5% acid in 3 gallons of solution). These rate schemas are the building blocks of algebraic problem models. The story text usually specifies one or two of the members of a schema (Unit_1, Unit_2, rate). If only one member is missing, it can be computed from the formula above. Algebra problems usually require the instantiation of more than one schema, together with the relation between them. Thus, we might have two kinds of nuts with different costs and mix the two according to certain specifications, requiring three cost-per-unit schemas. To construct a problem model, the problem solver must (a) pick the right schema, which can be done on the basis of textual cues, just as in the case of the arithmetic schemas; (b) specify as many elements for each schema as the text allows; and (c) find or infer from the text the relationships among the schemas used. Once a problem model has been constructed, an equation for the problem can be found by constraint propagation within the problem model.

The problem for the student in all this is to tell whether the formalization that has been constructed adequately represents the problem situation. The real world gives the formalist some feedback as to the adequacy of the formalization: Bridges last for centuries or they collapse; the software is perfect but does not do what the customer imagined. Experienced algebraic word-problem solvers provide their own feedback by checking their answers for reasonableness against their situation model. But that is exactly where students need help; otherwise they come up with negative values for the length of a board and similar absurdities (Paige & Simon, 1966). Because algebra students cannot operate in the real world and receive feedback from it, the next best thing is to supply them with a substitute world in the form of an animation that acts out what is implied by the problem model they have constructed. Nathan, Kintsch, and Young (1992) have built a tutor that shows the student what the implications of a proposed formalization are in the real world. It is not an intelligent tutor because the system does not understand the problem at all. It merely executes an animation as instructed by the student's problem model. (It paints fences, mixes solutions, stacks up piles of money, or lets objects move.) If the right event happens in the animation, the student knows that his solution was correct. If the wrong event happens, the student must figure out how to correct it. If nothing happens because the simulation was not given enough information to execute anything, the student realizes that one or more pieces are missing from his formal model and can try to complete it.

Suppose a student is trying to solve an overtake problem: A plane leaves Denver en route to Chicago with a certain speed; half an hour later, a second plane leaves on the same route with a greater speed. When will the second plane overtake the first plane? The student formulates his problem model and the tutor illustrates what happens according to that model: One plane flies away, then a second one follows and soon overtakes it; or, if the problem model is faulty, the second plane may never leave or may move with a slower speed, and the student realizes that he has done something wrong.

Problem solving often requires an explicit formal problem model. Helping students to construct such a model by providing feedback that allows them to use their everyday understanding of the problem situation can be an effective tutoring tool. The tutoring strategies sketched previously in the context of word problems in arithmetic and algebra have a potentially broad range of application. Making explicit the link between text, formalization, and action could be a powerful instructional tool that deserves further exploration. Problem solvers often understand their problem correctly in everyday terms but have trouble constructing the right formalization for it, which is necessary for the use of powerful, formal solution methods. If an animation can show them whether their problem model actually corresponds to the intended situation or not, they can use their correct situational understanding to repair their formal model.

Research on Reading Comprehension and the Teaching of Reading Comprehension

Research on reading has been an active field in the last few decades. Relatively large sums of money have been made available by federal agencies such as the National Institute of Mental Health (NIMH) to support this field. However, the focus of this research effort has been squarely on early reading instruction—on the study of decoding processes, rather than comprehension. To focus research on decoding was a perfectly defensible and successful strategy: We now have a fairly good understanding of the cognitive bases of decoding processes in reading and about reading instruction in the early grades. Surely, there remain problems to be resolved, but there exists an underlying consensus today about early reading instruction in the United States, as exemplified by the National Research Council report *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998). Educators know what to do, even if getting it done in schools on a national scale is still another matter.

There also is agreement among reading researchers today that research on reading comprehension lags far behind research on decoding processes and early reading instruction, and that it is time to shift the research focus onto reading comprehension beyond the early years. Recent assessments of research needs by the RAND Reading Study Group (RAND, 2002) and the Strategic Education Research Partnership report of the National Research Council's Panel on Learning and Instruction (Donovan, Wigdor, & Snow, 2003) agree on the need for a better understanding of the processes of text comprehension as well as instructional methods to improve comprehension.

My goal in this chapter has been to show that there exists a solid basis for further research in reading comprehension. We do not have to start from zero; there is a sparse but solid database, as well as a theoretical framework, that can serve at least as a good starting point for further research on reading comprehension. Throughout this chapter, open research questions have been pointed out, most pressingly about the formation of situation models and the modeling of macrostructures. There is much to be learned, but we also have already learned quite a bit about comprehension. This chapter was not intended as a general review of research on comprehension, but rather as a description of one particular research program and theoretical approach. A broader discussion would certainly

have provided further evidence of the considerable progress made in the study of reading comprehension in recent years.

The explicit goal of the comprehension research presented here is to inform instructional practice. As yet, this link is weak because there are so many unanswered questions and limited, conditional answers, but there is no reason to suppose that a focused research effort in this area would not yield results that achieve this goal.

Theories of discourse comprehension such as the one presented here are based on data from proficient readers. Indeed, these readers, as long as they read familiar material, can be considered to be comprehension experts. Comprehension for them is fluent, automatic, and easy. Well-established knowledge structures and skills are the basis for this automaticity. The goal of instruction is to help students become such expert readers. Paradoxically, however, comprehension instruction requires students to behave in very different ways than experienced readers. Because for student readers comprehension is not the automatic, fluent process that it is for mature readers, students need to engage in active problem solving, knowledge construction, self-explanation, and monitoring—activities very different from the automatic, fluent comprehension of experts. For the expert reader, comprehension is easy; to become an expert, comprehension must be hard work. Research on comprehension, therefore, has two quite distinct goals: (1) to describe expert comprehension with all its components and (2) to determine the training sequence that leads to this expert performance. What the student needs to do in training is quite different from how the expert operates. This is not a problem peculiar to comprehension training. Take, for example, ski instruction. Watching the instructor glide down a steep slope with elegant turns is not helpful to the novice skier. The novice must learn by doing things quite differently, and with much more effort, and the instructor must gradually, via a carefully thought out training sequence, bring the novice to the point where he can begin skiing like an expert, that is, when he is no longer a novice. Thus, if it is to be relevant for instruction, comprehension theory must pay attention not only to the final automatic comprehension that characterizes expert readers in familiar domains, but also to the strategies that support comprehension for the beginner, or for the expert who is faced with materials outside his or her domain of expertise.

Assessment plays a central role in gaining expertise in reading comprehension. This chapter has stressed how nontrivial comprehension assessment is. The levels of comprehension range from the superficial to the deep, from surface features to the textbase to the situation model. Assessing comprehension at these different levels is tricky because quite different tests are required. To teach comprehension, we need a thorough understanding of the different levels of comprehension and the tests that assess comprehension at these different levels. Richer comprehension tests need to be developed and evaluated that adequately assess the different aspects of comprehension. Furthermore, not only must teachers be able to tell how well students understood something, but the students themselves also must have tools to assess their comprehension or lack thereof. People are notoriously bad at this task, and one of the goals of comprehension research must be to find better, and more practical, ways to assess comprehension.

Research on comprehension will probably see a big boost in the next decades. To fulfill its potential it will have to find the right balance between observation, experiment, and theory. Careful studies of the basic cognitive processes in comprehension are needed, together with research on instructional practices and tools that support effective comprehension. Our goal should be a comprehensive theory of comprehension that allows us to understand how people, novices as well as experts, will react in novel situations. We cannot always perform a new experiment for every new question; instead, we need a broad theoretical framework that provides reasonably good answers to these questions. Educational researchers need a reliable theory to navigate by, much as engineers do in other fields, when they only occasionally resort to experiment because they know they can rely on their computations, except for special problems.

Notes

1 The term *proposition* was borrowed from logic, where it is used quite differently.

2 The distinction between textbase and situation model is made for the convenience of the theorist; mental representation integrates aspects of both.

3 Different word *meanings* are unrelated, as in *bank-(of river)* and *bank-(financial institution)*; different word senses are related, as in *chill-(bodily coldness with shivering)* and *chill-(moderate coldness)*.

4 For comparison, the auto-summary computed by MS Word is "Later, people built windmills to grind wheat and other grains."

5 Based in part on W. Kintsch (1998, chap. 6).

6 Based in part on W. Kintsch, Patel, & Ericsson (1999).

7 The term *long-term memory* is used broadly here; it includes personal experiences as well as general knowledge.

8 The section is based in part on W. Kintsch (1998, chap. 9).

9 Elaborative recall is that portion of a recall protocol that is left over when verbatim or paraphrased reproductions of the text are deleted.

References

Albrecht, J.E., & O'Brien, E.J. (1995). Goal processing and the maintenance of global coherence. In R.F. Lorch & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 263–278). Hillsdale, NJ: Erlbaum.

Albrecht, J.E., O'Brien, E.J., Mason, R.A., & Myers, J.L. (1995). The role of perspective in the accessibility of goals during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 364–372.

Battig, W.F. (1979). The flexibility of human memory. In L.S. Cermak & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 23–44). Hillsdale, NJ: Erlbaum.

Beck, I.L., McKeown, M.G., Sinatra, G.M., & Loxterman, J.A. (1991). Revising social studies texts from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*, 251–276.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147.

Bisanz, G.L., LaPorte, R.E., Vesonder, G.T., & Voss, J.F. (1978). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 17*, 337–357.

Bransford, J.D., Barclay, J.R., & Franks, J.J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209.

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

Britton, B.K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*(3), 329–345.

Chase, W.G., & Ericsson, K.A. (1981). Skilled memory. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

DeCorte, E., Verschaffel, L., & DeWinn, L. (1985). The influence of rewording verbal problems on children's problem representation and solutions. *Journal of Educational Psychology, 77*, 460–470.

Donovan, M.S., Wigdor, A.K., & Snow, C.E. (Eds.). (2003). *Strategic Education Research Partnership*. Washington, DC: National Academy Press.

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245.

Ferstl, E., & Kintsch, W. (1999). Learning from text: Structural knowledge assessment in the study of discourse comprehension. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 247–278). Mahwah, NJ: Erlbaum.

Fincher-Kiefer, R.H. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16*(1), 99–124.

Foertsch, J., & Gernsbacher, M.A. (1994). In search of complete comprehension: Getting "minimalists" to work. *Discourse Processes, 18*(3), 271–296.

Garrod, S., O'Brien, E.J., Morris, R.K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 250–257.

Glenberg, A.M., Kruley, P., & Langston, W.E. (1994). Analogical processes in comprehension: Simulation of a mental model. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 609–640). San Diego, CA: Academic Press.

Graesser, A.C. (1981). *Prose comprehension beyond the word*. New York: Springer.

Graesser, A.C., & Kreuz, R.J. (1993). A theory of inference generation during text comprehension. *Discourse Processes, 16*(1/2), 145–160.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395.

Graesser, A.C., & Zwaan, R.A. (1995). Inference generation and the construction of situation models. In C.A. Weaver, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 117–139). Hillsdale, NJ: Erlbaum.

Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episodic structure of algebra story problem solving. *Cognition and Instruction, 6*, 223–283.

Henley, N.M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8*, 176–184.

Johnson, C.S. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241–254.

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*(3), 418–439.

Kintsch, E., Steinhart, D., Stahl, G., Matthews, C., Lamb, R., & LSA Research Group. (2000). Developing summarization skills through the use of LSA-backed feedback. *Interactive Learning Environments, 8*(2), 87–109.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (2002). On the notion of theme and topic in psychological process models of text comprehension. In M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary studies* (pp. 157–170). Amsterdam: John Benjamins.

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 150–159.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.

Kintsch, W., Patel, V.L., & Ericsson, K.A. (1999). The role of long-term working memory in text comprehension. *Psychologia, 42*, 186–198.

Kintsch, W., & van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*, 363–394.

Landauer, T.K. (1998). Learning and representing verbal meaning: Latent Semantic Analysis theory. *Current Directions in Psychological Science, 7*, 161–164.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259–284.

Landauer, T.K., Laham, D., & Foltz, P.W. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems*, 27–31.

Le, E. (2002). Themes and hierarchical structures of written text. In M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary studies* (pp. 171–188). Amsterdam: John Benjamins.

Long, D.L., & Golding, J.M. (1993). Superordinate goal inferences: Are they automatically generated during reading? *Discourse Processes, 16*(1/2), 55–73.

Long, D.L., Golding, J.M., & Graesser, A.C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language, 31*, 634–647.

Long, D.L., Oppy, B.J., & Seely, M.R. (1994). Individual differences in the time course of differential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1456–1470.

Louwerse, M. (2002). Computational retrieval of texts. In M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary studies* (pp. 189–216). Amsterdam: John Benjamins.

Magliano, J.P., Baggett, W.B., Johnson, B.K., & Graesser, A.C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes, 16*(1/2), 35–53.

Mani, K., & Johnson-Laird, P.N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10*(2), 181–187.

Mannes, S.M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91–115.

Mayer, R.E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional Science, 10*, 135–175.

McDaniel, M.A., Blischak, D., & Einstein, G.I. (1995). Understanding the special mnemonic characteristics of fairy tales. In C.A. Weaver, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 157–176). Hillsdale, NJ: Erlbaum.

McKeown, M.G., Beck, I.L., Sinatra, G.M., & Loxterman, J.A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly, 27*, 78–93.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440–466.

McKoon, G., & Ratcliff, R. (1995). The minimalist hypothesis: Directions for research. In C.A. Weaver, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 97–116). Hillsdale, NJ: Erlbaum.

McNamara, D.S., & Healy, A.F. (1995). A generation advantage for multiplication skill and nonword vocabulary acquisition. In A.F. Healy & L.E. Bourne (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 132–169). Newbury Park, CA: Sage.

McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1–43.

McNamara, D.S., & Kintsch, W. (1996). Learning from text: Effect of prior knowledge and text coherence. *Discourse Processes, 22*, 247–288.

Miller, J.R., & Kintsch, W. (1980). Readability and recall for short passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 335–354.

Moravcsik, J.E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology, 47*(2), 360–374.

Myers, J.L., O'Brien, E.J., Albrecht, J.E., & Mason, R.A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 876–886.

Myers, J.L., Shinjo, M., & Duffy, S.A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26*, 453–465.

Nathan, M.J., Kintsch, W., & Young, E. (1992). A theory of word algebra problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*, 329–389.

O'Brien, E.J. (1995). Automatic components of discourse comprehension. In R.F. Lorch & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 159–176). Hillsdale, NJ: Erlbaum.

O'Brien, E.J., Shank, D.M., Myers, J.L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 410–420.

Olson, J.R., & Biolsi, K.J. (1991). Techniques for representing expert knowledge. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 240–285). Cambridge, UK: Cambridge University Press.

Paige, J.M., & Simon, H.A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory*. New York: Wiley.

Perfetti, C.A. (1993). Why inferences might be restricted. *Discourse Processes, 16*(1/2), 181–192.

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language, 24*, 503–518.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Rayner, K., Pacht, J.M., & Duffy, S.A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33*, 527–544.

Rips, L.J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.

Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437–442.

Schank, R.C., & Abelson, R.P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Schmalhofer, F., McDaniel, M.A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*, 105–132.

Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Schvaneveldt, R.W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood NJ: Ablex.

Singer, M., Graesser, A.C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language, 33*, 421–441.

Snow, C.E., Burns, M.S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.

Till, R.E., Mross, E.F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16*(4), 283–298.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes, 16*(1/2), 3–34.

van den Broek, P.W. (1990). Causal inferences in the comprehension of narrative texts. In A.C. Graesser & G.H. Bower (Eds.), *Inferences and text comprehension* (Psychology of learning and motivation, Vol. 25, pp. 175–194). San Diego: Academic Press.

van der Meer, E., Beyer, R., Heinze, B., & Badel, I. (2002). Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 770–779.

van Dijk, T.A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Erlbaum.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Voss, J.F., & Silfies, L.N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction, 14*(1), 45–68.
Wade-Stein, D., & Kintsch, E. (2003). *Summary Street: Interactive computer support for writing.* Manuscript submitted for publication.
Wolfe, M.B., Schreiner, M.E., Rehder, R., Laham, D., Foltz, P.W., Landauer, T.K., et al. (1998). Learning from text: Matching reader and text by Latent Semantic Analysis. *Discourse Processes, 25*(2/3), 309–336.
Yekovich, F.R., Walker, C.H., Ogle, L.T., & Thompson, M.A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. In A.C. Graesser & G.H. Bower (Eds.), *Inferences and text comprehension* (Psychology of learning and motivation, Vol. 25, pp. 259–278). New York: Academic Press.
Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 386–397.
Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185.

47

A Dual Coding Theoretical Model of Reading

Mark Sadoski and Allan Paivio

Dual Coding Theory (DCT) is an established theory of general cognition that has been directly applied to literacy. This theory was originally developed to account for verbal and nonverbal influences on memory, and it has been extended to many other areas of cognition through a systematic program of research over many years (Paivio, 1971, 1986, 1991). DCT has been extended to literacy as an account of reading comprehension (Sadoski & Paivio, 1994; Sadoski, Paivio, & Goetz, 1991), as an account of written composition (Sadoski, 1992), and as a unified theory of reading and writing (Sadoski & Paivio, 2001). For the fullest understanding of the theory, these references and the specific studies they cite should be consulted. This article briefly discusses the DCT account of certain basic processes in reading, including decoding, comprehension, and response.

The value of explaining reading under the aegis of a theory of general cognition is compelling. Reading is a cognitive act, but there is nothing about reading that does not occur in other cognitive acts that do not involve reading. We perceive, recognize, interpret, comprehend, appreciate, and remember information that is not in text form as well as information that is in text form. Cognition in reading is a special case of general cognition that involves written language. Theories specific to reading must eventually conform to broader theories of general cognition for scientific progress to advance. DCT provides one vehicle for that advancement.

Another value offered by DCT is that it provides a combined account of decoding, comprehension, and response. Theories of reading often focus on one or another of these aspects of reading but not all. As we shall see, the same basic DCT principles apply to grapheme–phoneme correspondences, word meaning, grammar, the construction of mental models of text episodes, and even imaginative responses to text. In this article, we will briefly explain the theory's basic assumptions; provide accounts of decoding, comprehension, and response; compare and contrast DCT with other theories of reading; and discuss its implications for research and practice.