12

Comprehension

Walter Kintsch and Katherine A. Rawson

"Comprehension" refers to both a set of empirical phenomena and a theoretical construct. The phenomena are ill defined, because the concept of comprehension that psychologists took over from everyday language use is fuzzy, as such concepts are. In practice, however, most psychological research on comprehension has dealt with discourse comprehension. In discourse comprehension for skilled adult readers, analytical reasoning is required only when the normal course of understanding breaks down: normal reading or listening is more akin to perception than to problem solving. As a theoretical construct, therefore, "comprehension" contrasts with conscious, deliberate problem solving. Perception and comprehension can both be described as spontaneous constraint satisfaction processes. Consider, for example, the classical sentence pair:

The turtle sat on a log. A fish swam under the log.

from which readers immediately conclude that

The turtle was above the fish.

It is possible to arrive at this conclusion through a reasoning process, using rules like

If A is above B, and C is under B
then A is above C.

Alternatively, a reader who constructs a mental image of this scene can immediately "see" that the turtle is above the fish. Most people can understand the simple sentence pair above without needing to reason it out analytically, because they know about turtles and logs and water and fish, combining these constraints into a readily comprehended scene (though, of course, analytic problem solving may be required in more complex situations). Kintsch (1998) has therefore argued that comprehension forms a contrasting paradigm for cognition, complementing but by no means supplanting problem solving as conceived in the classical monograph of Newell and Simon (1972).

Processes Underlying Text Comprehension

Text analysis has long been the province of linguists and logicians. When psychologists became interested in text comprehension, they first followed this lead, but soon developed their own agenda: their interest was in text processing, rather than the analysis of the text per se. Below, we discuss some of the processes involved in text comprehension along with some of the major issues that have arisen in trying to understand how people understand text. We will begin by outlining the basic processes involved in text comprehension and will then illustrate them with a simple example. Subsequently, we will examine a few of these processes in greater detail.

Levels of processing

Text comprehension is often described as involving processing at different levels (Kintsch, 1998): In keeping with this convention, we use "levels" terminology herein (e.g., "textbase," "microstructure," "macrostructure," and "situation model," described further below) to provide the reader with convenient and conventional terms used by researchers to describe the different kinds of information that may be represented by comprehension processes, although we do not claim that these kinds of information are necessarily stored in distinct or separate representations.

- *First, there is the linguistic level, or processing of the particular words and phrases contained in the text itself.* The reader must decode the graphic symbols on a page. Perceptual processes are involved, as well as word recognition and parsing (the assignment of words to their roles in sentences and phrases). Treatment of many of these processes is available in other chapters, and thus we will focus subsequent discussion on the higher-level processes involved in comprehension introduced below.
- *Semantic analysis determines the meaning of the text.* Word meanings must be combined in ways stipulated by the text, forming idea units or propositions. Propositions are interrelated in a complex network, called the *microstructure* of the text. One of the main dimensions along which propositions can be related to one another is coreference, which occurs when two or more propositions refer to the same concept (also referred to as "argument overlap"). Psychologically, the microstructure is constructed by forming propositional units according to the words of the text and their syntactic relationships and by analyzing the coherence relations among these propositions, which are often, but not always, signaled by cohesion markers at the linguistic level. Inferences, such as simple bridging inferences or pronoun identification, are often necessary to arrive at a coherent microstructure. Several models of microstructure for-

mation have been proposed, such as the construction-integration (CI) model of Kintsch (1988, 1998), the landscape model (van den Broek, Risden, Fletcher, & Thurlow, 1996) and the Langston, Trabasso, and Magliano (1998) model.

However, there is more to the meaning of a text than word meanings and the interrelationships between propositions. Whole sections of a text are also related semantically in specific ways. That is, the microstructure itself is organized into higher-order units. This global structure of a text is called the *macrostructure*. Macrostructure formation involves the recognition of global topics and their interrelationships, which are frequently conventionalized according to familiar rhetorical schemata (see Le, 2002, for demonstration that macrostructure formation can be modeled within the framework of the CI model). Microstructure and macrostructure together are called the *textbase*.

- *The textbase represents the meaning of the text, as it is actually expressed by the text.* But if a reader only comprehends what is explicitly expressed in a text, comprehension will be shallow, sufficient perhaps to reproduce the text, but not for deeper understanding. For that, the text content must be used to construct a *situation model*; that is, a mental model of the situation described by the text. Generally, this requires the integration of information provided by the text with relevant prior knowledge and the goals of the comprehender. A model in the tradition of the CI model that focuses on the situation level rather than the propositional textbase has been investigated by Louwerse (2002). Similarly, Schmalhofer, McDaniel, and Keefe (2002) have extended the CI model so that it explicitly accounts for the interaction of the propositional level and the situation model level. One important fact to note about the process of constructing situation models is that it is not restricted to the verbal domain. It frequently involves imagery, emotions, and personal experiences.

We illustrate these processes with a simple example. The text is an instructional text about the functioning of the heart, entitled "The Circulatory System." In figure 12.1, the macrostructure of the text is diagrammed: it is signaled explicitly in the text by five subheadings; the last section, which is quite long, is further subdivided into three topical subsections.

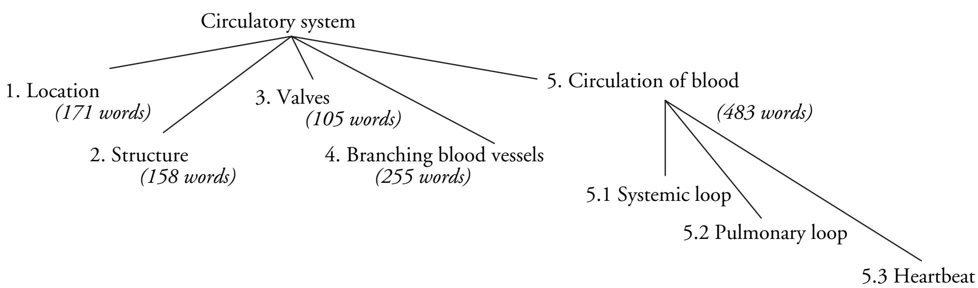

Figure 12.1 A diagram of the macrostructure of an instructional text about the functioning of the human heart.

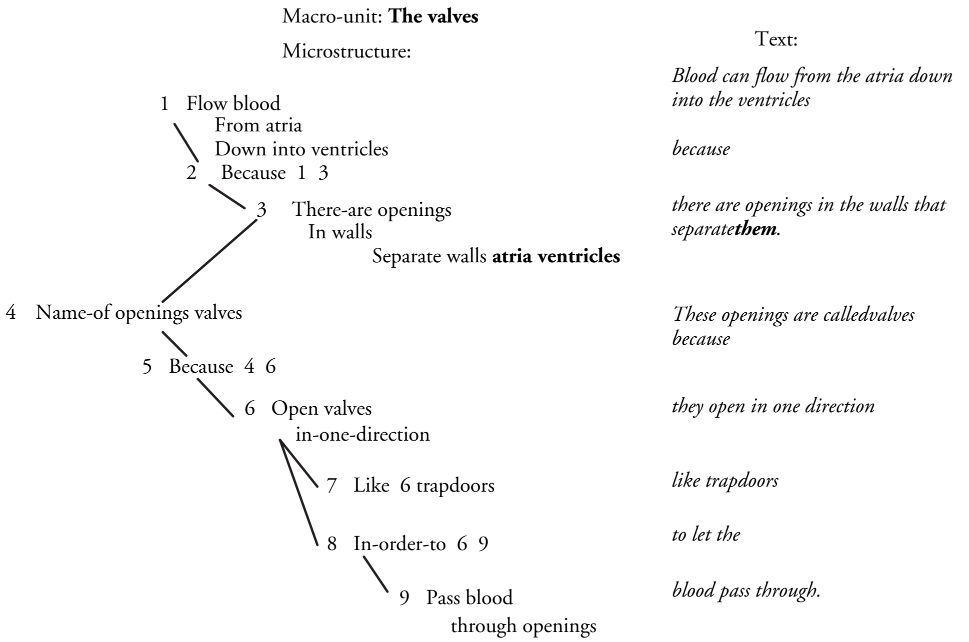

Figure 12.2 A sample of the microstructure based on two sentences from "The Valves" section of the instructional text on the heart.

Figure 12.2 shows an example of the microstructure of the text, specifically, the first two sentences of the macro-unit "The Valves." Note the hierarchical organization of the microstructure with the proposition introducing the concept VALVES at the top of the hierarchy. Not all relations among propositions are shown in the graph (e.g., the repetition of BLOOD in Propositions 1 and 9). The construction of the microstructure very closely follows the linguistic structure of the sentences. One relation is left implicit in the text, such that an inferential process is required to identify the referents of the pronoun "them" in the first sentence (this process is referred to as "anaphor resolution," which we will discuss further below).

The situation model constructed on the basis of these two sentences is shown in figure 12.3. This situation model is a diagrammatic representation of the human heart, showing that it is divided into two halves, and each half comprises an upper chamber (the atrium) separated by a valve from the lower chamber (the ventricle). The situation model contains much more than the information expressed in the two sentences under discussion. To understand these sentences, the reader must retrieve prior knowledge about the heart and integrate it with the new information provided by the text. The prior information (which for some readers might be prior knowledge about the heart but in this case is also information presented earlier in the text) is depicted with dotted lines: it involves the spatial layout of the chambers of the heart and their corresponding names. A reader who does not retrieve this information from his or her memory and integrate it with the new information provided by the text does not really understand the text, even if he or she formed a correct textbase. Such a reader might be able to reproduce the text by rote, or

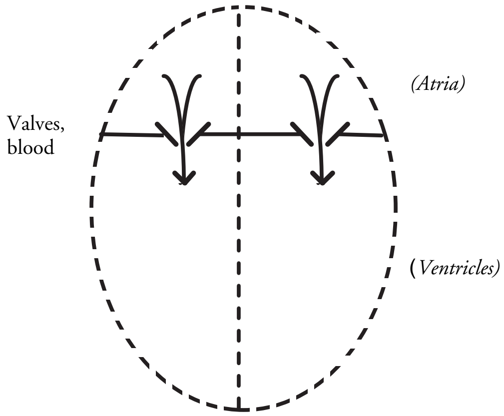

Figure 12.3 A depiction of information that could be represented in the situation model based on the two sentences from "The Valves" section presented in figure 12.2. Information that must be drawn from prior knowledge is indicated with dotted lines or parentheses; information that is explicitly stated in the sentences is indicated with solid lines.

even make some semantic judgments: for instance, on the basis of the hierarchical textbase shown in figure 12.2, the reader could say that the sentences were about "valves." However, this reader could not make correct inferences about the functioning of the heart.

With this overview and intuitive example in hand, we now examine further some of the processes involved in construction of the textbase and the situation model. Of course, there are many more processes involved than can be discussed here, and so our discussion is limited to an illustrative sample of the many component processes that are considered fundamental to text comprehension.

Before we begin, a brief discussion of some of the performance measures used to assess comprehension processes may be helpful for some readers. In general, the various measures used are of two basic kinds, *on-line* and *off-line*. On-line measures are those that are taken while a participant is reading. One widely used on-line measure is reading time, or recording of the time spent reading a particular region of text (which may range from a word to a phrase or sentence to an entire section of text). Other on-line measures include speeded response tasks that are performed during reading, such as lexical decision (i.e., deciding as quickly as possible whether a string of letters forms an English word), naming (i.e., pronouncing a visually presented word as quickly as possible), and recognition (i.e., deciding as quickly as possible whether a word appeared earlier in the text or not). Response times from these kinds of tasks are often used to infer the relative activation of various concepts or inferences. Off-line measures are those that are taken after a participant is done reading. Off-line measures come in many forms, although they usually involve some form of test question (e.g., free recall, short answer, multiple choice, problem solving). The kinds of questions typically used vary in the extent to which they tap primarily memory for the text material, deeper understanding of the content, or both.

Both on-line and off-line measures have advantages and disadvantages. For example, because on-line measures are temporally closer to the actual processing of the text, they tend to support more fine-grained inferences about the nature of the underlying processes. However, many of the on-line measures are somewhat disruptive and may encourage task-specific strategies that may not reflect "normal" reading processes. Additionally, some of the on-line measures may reflect transient activation of conceptual information rather than the more permanent inclusion of that information into the final representation. In contrast, off-line measures do not disrupt the reading process, and they are more indicative of the more lasting representational outcome of those processes. However, off-line measures tend to be less informative with respect to specific aspects of the way in which particular reading processes operate, and they are also subject to reconstructive processes at time of test or loss due to forgetting. Ultimately, text comprehension – both the underlying processes and the representational outcomes of those processes – will best be understood with converging evidence from multiple different measures.

Textbase Formation

As mentioned above, an important aspect of constructing the textbase involves combining word meanings to form propositions and then establishing the interrelationships between these propositions to generate the microstructure of the text. There are several different dimensions along which propositions may be related. For example, the content of two or more propositions may be related because of logical implication, because of cause and effect relationships, or because of reference to the same entity or event. This latter form of relation is referred to as *coreference* (or argument overlap), which is perhaps the most common dimension along which propositions are assumed to be related to one another. Below, we examine the factors that influence how readers establish coreference to illustrate one way in which propositions may be interrelated.

As was also mentioned previously, a textbase consists of a macrostructure as well as the microstructure formed by establishing connections between propositions. In relative terms, much less research has focused on macrostructure formation than on microstructure formation. However, one area of research relevant to understanding macrostructure formation has explored how readers identify the higher-level ideas or topics around which the microlevel units are organized and interrelated, which we discuss further below.

Establishing coreference: anaphor resolution

Often, coreference is made explicit in the text, when the same words are used to refer to the same concepts (e.g., "*Dan* backed his *car* out of the garage. Unfortunately, *Dan* drove the *car* right over his brother's bike"). In many cases, however, the referential relationships are not explicit, such as when pronouns are used to refer back to a previously mentioned concept or when different terms are used to describe the same entities (e.g., "*Dan* backed his *car* out of the garage. Unfortunately, *he* drove the *old sedan* right over his brother's bike"). Any linguistic device that can be used to refer back to a previously mentioned concept is called an *anaphor*, and *anaphor resolution* is the psychological process of identifying the previously mentioned concept (or *referent*) to which an anaphor refers. To resolve an anaphor, one or more candidate referents must be activated in working memory and the appropriate referent must be selected for integration with the current content. An anaphor may go unresolved if no referent is active, or if no one referent is sufficiently salient to warrant selection when multiple candidates are active (e.g., Greene, McKoon, & Ratcliff, 1992). Thus, it is important to understand what factors can influence referent activation and referent selection.

As illustrated in the examples above, several different kinds of anaphor exist, including pronouns, synonyms, and repeated nouns. These various kinds of anaphor tend to serve different functions and thus have different distributional patterns in the language. Pronouns are typically used to refer to recently mentioned or focused concepts that have been explicitly introduced in the text and that are currently activated in working memory. Highly explicit anaphora are much less likely to refer to recent or salient entities, and in fact, repeated noun anaphora have been shown to slow down processing when they do (e.g., Gordon & Chan, 1995). For example, after the sentence "Bill bought a car," readers are slower to read, "Bill drove it home" than they are to read, "He drove it home." Instead of salient entities, explicit anaphora are much more likely to refer back to more distant or less salient concepts that are likely no longer active, and they also often refer to concepts that have not been explicitly mentioned. Thus, the form of the anaphor itself can provide some signaling information about where or what the appropriate referent may be.

Other sources of linguistic information also contribute to anaphor resolution. One factor is the morphosyntactic features of pronouns, such as gender and number (e.g., Carreiras, Garnham, & Oakhill, 1996), although the use of gender information may not be as general as intuition would suggest (for discussion, see McDonald & MacWhinney, 1995). Other linguistic sources of information that can provide information about the appropriate referents for anaphora include structural factors, including various syntactic constraints (e.g., whether a potential referent participates in the same clause as the anaphor; whether the referent is in a syntactically prominent position), order of mention within a sentence (e.g., Gernsbacher, 1990), and structural parallelism. According to the heuristic of structural parallelism, a potential referent that holds the same grammatical role as the anaphor is preferred over one that does not. As one demonstration of the operation of this heuristic in anaphor resolution, Chambers and Smyth (1998) presented readers with short texts in which a context sentence was followed by a sentence containing an ambiguous pronoun in either the subject or object position. For example, (1) was followed by either (2) or (3):

(1) *Leonard handed Michael a sandwich.*
(2) *Then he passed Carla an apple.*
(3) *Then Carla passed him an apple.*

Readers were much more likely to resolve the pronoun *he* in (2) as referring to *Leonard*, whereas they were much more likely to resolve the pronoun *him* in (3) as referring to *Michael*.

In addition to linguistic factors, various semantic factors can also influence anaphor resolution. One widely researched semantic factor concerns the *implicit causality* of verbs, which refers to "a property of transitive verbs in which one or the other of the verb's arguments is implicated as the underlying cause of the action or attitude" (Long & De Ley, 2000, p. 546). For example, the verb *question* implicates the sentence's subject as the cause of the questioning, as in (4), whereas the verb *praise* implicates the sentence's object as the cause of the praising, as in (5).

(4) *John questioned Chris because he wanted the correct answers.*
(5) *John praised Chris because he knew the correct answers.*

The influence of implicit causality on anaphor resolution is apparent in the bias of readers to prefer the causally implicated entity as the referent of an anaphor (e.g., Stewart, Pickering, & Sanford, 2000).

In addition to the semantic information associated with particular words such as verbs or connectives, semantic information at the discourse level can also influence anaphor resolution. For example, contextual information can establish pragmatic plausibility that favors one potential referent over another. Upon reading

(6) *Scott stood watching while Henry fell down some stairs. He ran for a doctor.*

Scott is a much more plausible referent for *he* than is *Henry*, given that someone who just fell down the stairs is unlikely to be able to run anywhere. Indeed, various studies have shown that pragmatic plausibility as well as other forms of discourse-level semantic information (e.g., discourse focus) can influence anaphor resolution (e.g., Gordon & Scearce, 1995; Stevenson, Knott, Oberlander, & McDonald, 2000).

The various sources of information that can be used to identify the referent for an anaphor will be more or less relevant to resolving different kinds of anaphora, because they are more or less relevant given the situations in which those different anaphora tend to be used. For instance, to the extent that more explicit anaphora are used to refer back to distant entities, structural parallelism is unlikely to exert much influence on resolution of these anaphora because surface-level information about the referent's sentence structure will have been lost during the processing of intervening sentences. In contrast, structural factors are more likely to factor in to the resolution of pronominal anaphora, as pronouns tend to occur in the same local context as their referent. Likewise, the degree of semantic overlap between the anaphor and the referent is likely to play an important role in the resolution of many noun anaphora that require the reactivation of the referent entity, to the extent that such overlap serves as a retrieval cue. However, this factor is much less relevant for resolving pronouns because they are semantically impoverished and their referent is likely to still be active upon their encounter.

Although the influence of a given factor on anaphor resolution may depend upon the form of anaphor to be resolved, one proposal is that the same basic constraint-satisfaction process underlies anaphor resolution in all cases. Current research that focuses on how these various factors interact and the time course of their respective contributions to anaphor resolution (e.g., the extent to which a factor influences activation, selection, or both) promises to shed light on this issue.

Macrostructure formation: topic identification

Establishing connections between propositions that are referentially, causally, logically, or otherwise related is one important step in constructing the textbase. However, another important step is to represent the macrostructure of the text, which involves relating larger units of text into a topical structure. One key aspect of this process involves identifying the important themes or topics in a text. In some cases, a text may contain *signaling devices* that explicitly indicate topical information, such as when the theme of a text is explicitly provided by an appropriate title or when the topics of a text are explicitly indicated by outlines, summaries, or section headers (as was the case in the instructional text on the human heart described earlier). Such signaling devices have been shown to improve subsequent recall for the topics discussed within a text as well as organization of those topics (e.g., Lorch, Lorch, & Inman, 1993).

But many texts will not contain signaling devices that explicitly indicate the topical structure of the text, and research has shown that various factors will influence the content that readers identify as the important topical information in these cases. Many of these factors are surface cues, such as typeface, rhetorical cues, repetition of concept words, or structural features of the text. For example, Surber (2001) demonstrated that the degree of repetition of a concept label can influence the extent to which readers identify that concept as an important topic. Concepts that appeared frequently in the text were either referred to using the same concept label each time or were referred to using nonrepeating labels (e.g., paraphrases). Additionally, half of the target concepts were actually important topics and half were unimportant topics. After reading the text, participants were given a topic recall test in which they were asked to list the main topics and subtopics. Target concepts were listed more often in the repetition condition than in the nonrepetition condition, regardless of the actual importance of the information.

As a further example of the influence of surface cues on topic identification, Budd, Whitney, and Turley (1995) note that readers often assume that the first sentence of a text or paragraph contains important topical or thematic information. Indeed, reading times for initial sentences can be 500–1,000 ms longer than for comparable sentences. Budd et al. presented readers with short texts that began with a topic sentence followed by several sentences that provided supporting details. In one condition, the texts were presented intact. In a second condition, the topic sentence was removed such that the first sentence of the paragraph was one that provided detail information. Reading times for the detail sentence were significantly longer when it was the first sentence of the text than when it was the second sentence of the text (i.e., when the topic sentence was presented) and were just as long as reading times for the topic sentences. In a similar study, Kieras (1980) had participants read short texts that each had a target sentence expressing the intended theme of the text. The target sentence either appeared as the first sentence of the text or was embedded in the middle of the text. Immediately after reading a text, participants were asked to write a short sentence that described the main idea or theme of the text. The responses were closer to the target sentence in content when the target appeared at the beginning of the text rather than in the middle. Likewise, the responses were more similar in content to the alternate first sentence when it appeared first rather than in the middle, even though it did not express main thematic information. Kieras found similar results in a second experiment using texts that each contained two related topics that could be discussed in either order. Participants were then given a two-alternative forced choice task for each text in which they were to decide which of two statements was a better title for the text (each title only mentioning one of the two topics in the text). Readers' choices were heavily influenced by which topic had been discussed first in the text.

Kieras also showed that for a subset of the texts in which one topic was conceptually superordinate to the other (e.g., whole–part relationship), readers' choices were also biased to favor the superordinate topic over the subordinate topic overall, although the strength of the bias was weaker when inconsistent with order of mention. Nonetheless, this superordinate bias demonstrates a role for semantic factors in addition to surface-level cues in identifying important topics or themes. Several other studies have subsequently established the role of semantic content in topic identification (e.g., Goldman, Saul, & Coté, 1995; Hyönä, 1994). For example, Hyönä (1994) showed longer reading times on sentences that introduced new topics (relative to topic continuation sentences equated along important characteristics such as length and syntactic structure), even when using a sentence-by-sentence presentation mode that did not denote paragraph boundaries.

Finally, topic identification may be influenced by a reader's prior knowledge. Relevant prior knowledge may include knowledge about the typical structure of texts within a domain or text genre, although this sort of knowledge may be more important for narratives (most of which have similar discourse structures) than for expository texts. Readers can also rely upon knowledge about content within the domain. For example, Dee-Lucas and Larkin (1988) presented beginning learners and advanced learners in physics (two to three undergraduate physics courses versus graduate-level training in physics, respectively) with instructional texts containing target sentences in one of two versions. In both versions, the substantive content of the target sentence was the same, but the content was presented either as a definition or as a fact. After reading the text, individuals were asked to rate the importance of each sentence in the text. The beginning learners rated the target sentences as more important when presented as definitions than when presented as facts, suggesting that they had developed a rudimentary understanding that definitions are generally more important than facts. In contrast, the ratings of the advanced learners did not differ with version, presumably because they could assess the importance of the information based on its content rather than having to rely on a less discriminating, category-membership heuristic.

In sum, various sources of information can influence topic identification, an important component in the process of representing the macrostructure in a textbase. Of course, topic identification is not the only macrolevel process involved in textbase construction. However, much less research has explored macrolevel processes than microlevel processes, and thus further investigating how macrostructures are represented will be an important issue for future research.

The Situation Model

As outlined earlier, the textbase represents the meaning actually expressed by the text. But deeper comprehension also depends upon the construction of a *situation model*, or the representation of the situation described by the text. This is primarily achieved via the integration of information provided by the text with relevant prior knowledge. Thus, inferencing is critically involved in forming a situation model.

Inferences

Inferences are necessary in constructing the textbase (at both the micro and macro levels), and they play a crucial role in forming a coherent situation model. Texts are almost never fully explicit, so there are always gaps left to be filled in by the reader. The gaps may be local as in

(7) *Fred parked the car. He locked the door*,

where the reader must realize that the *door* is the

(8) *car door*.

Or the gaps may be global, as when the theme of a story is not explicitly stated and left for the reader to construct, or when a reader must realize that a particular paragraph in an essay provides an example for a point made earlier. This gap filling has traditionally been labeled "inference." This is a somewhat unfortunate terminology because it lumps together processes that are quite distinct psychologically and that differ dramatically in the demands they make on the reader (Kintsch, 1993, 1998). First, inferences vary along a dimension from automatic to controlled. Automatic inferences are made quickly and easily, such as the bridging inference linking *car* and *door* in (7). Controlled processes, on the other hand, can be highly resource demanding, as is the case, for instance, if a text requires syllogistic reasoning. A second dimension along which inferences in text comprehension vary is whether they are knowledge-based or text-based. Example (7) involves a typical knowledge-based inference – the reader knows that cars have doors. In contrast, if we conclude from

(9) *Fred is taller than Mary and Mary is taller then Tim*.

that

(10) *Fred is taller than Tim*,

we are not using what we already knew about *Fred* and *Tim*, but must employ the information provided by the text. In the literature, all these psychologically rather distinct processes are called "inferences" – but it is crucial that we keep in mind the important differences between various types of inference in text comprehension. (And it almost goes without saying that most of these inferences have nothing to do with what logicians call inference.)

The principal question about inferences in text comprehension has been what types of inferences are made and when they are made. On the one hand, one can focus on all the inferences that can plausibly be made (e.g., Graesser, 1981). Of course, not all readers will actually generate all inferences at all times. So what do readers actually do? This simple question turns out not to have a simple answer. Whether or not a reader draws a particular inference depends on a great many factors. Under some conditions, readers are minimalists (McKoon & Ratcliff, 1992), making only those inferences that are absolutely needed to understand the text. For instance, (7) cannot be understood without making the bridging inference linking *car* and *door*; but (9) can be understood perfectly well, and many readers will not bother to figure out the relationship between Fred and Tim, unless asked to do so. On the other hand, it is easy to find conditions under which readers are far more active than minimally required (Graesser, Singer, & Trabasso, 1994). For instance, readers typically infer causal antecedents but not causal consequences (Magliano, Baggett, Johnson, & Graesser, 1993). Given

(11) *The clouds gathered quickly and it became ominously dark.*
(12) *The downpour lasted only 10 minutes.*

readers infer

(13) *The clouds caused rain.*

But given only (11) readers do not necessarily jump to the conclusion (13). However, under the right conditions, it is quite possible to get readers to make predictive inferences (Klin, Guzmán, & Levine, 1999; Klin, Murray, Levine, & Guzmán, 1999).

What happens when a reader "makes the inferences" illustrated above? Is it a slow active process or a rapid, automatic one? Often, it is just that a piece of relevant knowledge is retrieved, such as that *cars* have *doors* or that *dark clouds* bring *rain*. It is not an active, controlled, effortful process, but rather happens quickly and relatively automatically. Kintsch (1998) argued that retrieval structures in long-term memory rapidly make available the relevant knowledge when the reader reads (7) or (11); similarly, Myers and O'Brien (1998) talk about a resonance process that provides the link between *car* and *door*, or *cloud* and *rain*. The fact that antecedent causal inferences are more likely than predictive inferences does not mean that the knowledge does not become available in the latter case, only that it is more likely to be used when it is linked to two items – (11) and (12) in the example above – than when it is only linked to one (Kintsch, 1998; Schmalhofer, McDaniel, & Keefe, 2002).

Many inferences in text comprehension are straightforward cases of knowledge activation. Automatic knowledge activation will work well as long as the text is in a highly familiar domain. However, the importance of active, controlled, constructive inferencing in comprehension can hardly be overestimated, especially with expository texts in less familiar topic domains. When we read a text in order to learn from it, by definition we are no longer on the kind of highly familiar ground where we can rely on retrieval structures to activate relevant knowledge. It is still necessary, however, to retrieve whatever relevant prior knowledge and experience we have, which can be a very effortful process, requiring conscious control. Without this effort, no learning is possible – the textual information will remain inert knowledge at best, not linked up with existing knowledge structures, and hence unusable.

Resource demanding, controlled inference processes in text comprehension are not restricted to knowledge retrieval. The situation model for a literary text may require construction at more than one level of analysis; to understand a story, the reader may have to infer the protagonists' motivations; to understand an argument, the exact relations between its components may have to be analyzed. Deep understanding always goes beyond the text in non-trivial ways, requiring the construction of meaning, not just passive absorption of information.

Knowledge representation

Comprehension requires inferences, and inferences require knowledge. Hence to understand text comprehension we must be able to understand how knowledge is used and how it is represented. Most psychological (and linguistic and artificial intelligence) models of knowledge representation are toy models that cannot deal with the sheer amount of information in human knowledge. Recently, however, Latent Semantic Analysis (LSA) has become available, which provides psychologists at least with a tolerably good approximation to human knowledge representation (Landauer, 1998; Landauer & Dumais, 1997; see also the website at http://lsa.colorado.edu).

LSA is a machine learning method that constructs a geometric representation of meaning that resembles the structure of human knowledge about words and texts. It constructs this representation simply from observing the contexts in which words are used in a large number of texts. Formally, the problem faced by LSA might be characterized by an equation that expresses the meaning of a document as a function of its words, their order, and interrelationships, as well as the (verbal and nonverbal) context (Landauer, 2002):

(13) $$meaning(document) = f\{word_1, word_2, word_3, \ldots\ word_n, context\}$$

To solve this equation (for a large number of documents), LSA makes some simplifying assumptions:

(14)
$$meaning(document_1) = meaning(word_{11}) + meaning(word_{12}) + \ldots\ meaning(word_{1n})$$
$$meaning(document_2) = meaning(word_{21}) + meaning(word_{22}) + \ldots\ meaning(word_{2m})$$
$$\vdots$$
$$meaning(document_k) = meaning(word_{k1}) + meaning(word_{k2}) + \ldots\ meaning(word_{1kz})$$

In other words, we disregard word order, syntax, as well as all context, and take the meaning of a document to be just the meaning of a bag of words. These are drastic simplifications, but, as we will see, enough information is retained in this way to produce useful results.

Consider a corpus of 11 M word tokens, 90 K word types, and 40 K documents, consisting of texts a typical high-school graduate might have read. This corpus clearly underspecifies word meanings, and is furthermore inconsistent. What LSA does is to extract from such a corpus a semantic representation that does not attempt to specify "the meaning" of each word and document in absolute terms (like a dictionary or encyclopedia would), but determines only the relations among all the words and documents. That is, LSA defines meaning as the relationship of a word (or document) to all other words and documents in the corpus. LSA does this by constructing a high-dimensional semantic space, using a standard mathematical technique called singular value decomposition for optimal dimension reduction to eliminate noise in the data.

Semantic relatedness in the LSA space is measured by the cosine between vectors representing words or documents, a statistic much like the familiar correlation coefficient. The cosine between randomly chosen words is .02 +/− .04. Below are some examples that show that the similarity measure calculated by LSA yields results not unlike human intuition:

doctor–doctors .79
doctor–physician .61
go–went .71
good–bad .65
she–her .98
blackbir–bird .46
blackbir–black .04
telephone– shark .01
telephone–justice .01

Note that in the original corpus, co-occurrence of the words *doctor* and *doctors* (or between singular and plural, in general) is quite low, because when one talks about a singular entity one rarely also mentions the plural, and vice versa. LSA, however, has inferred that these singulars and plurals are quite similar in meaning (not identical, though), because singular and plural forms tend to occur in similar contexts.

Many words have more than one meaning, and most have several senses, depending on context. In LSA, meaning is context free, but note that in a high-dimensional space, complex relationships can be naturally represented. For instance, the homonym *mint* has (at least) three senses, as in *leaves of a plant*, *flavored candy*, and *coin money*. The cosines between the word *mint* and these three phrases are .20, .23, and .33, respectively. Thus, *mint* is strongly related to each of these phrases that involve different meanings, even though these three phrases are not related to each other (the average cosine between the three phrases is only .05; for a more detailed treatment of polysemy in LSA, see Kintsch, 2001).

LSA is not restricted to computing the semantic similarities among words. Sentences and whole texts can be represented in the same semantic space, and hence can be readily compared with each other. The similarity measures that LSA computes for texts correlate well with human judgments, as is most dramatically shown by the ability of LSA-based systems to grade essays as well as expert human graders do (Landauer, Laham, & Foltz, 2000). LSA is not, however, a model of human comprehension processes; it is simply a representational system that allows researchers to represent the meaning of words and texts in such a way that the relations among the words and texts represented in LSA closely resemble human semantic judgments. It thus opens up numerous exciting possibilities for research and applications, only a few of which have so far been explored.

Situational dimensions

As discussed in the sections above, constructing a situation model involves inferences that usually depend upon general world knowledge. These inferences are often critical for representing the relationships between the entities and events described in the text. Zwaan and Radvansky (1998) have reviewed research suggesting that situation models can include representation of relationships along several different dimensions. Among other possible dimensions, research shows that situation models can include representations of the spatial locations of entities and events, the temporal relations between events, the causal relationships between actions or events, the goals and motives of protagonists, and the characteristics of protagonists and other important entities.

In much of the previous research, each of these dimensions has been examined in isolation. However, motivated by the observation that many of the situational dimensions are interrelated and that a full understanding of many texts will require representation of multiple dimensions, more recent research has begun investigating two or more dimensions at a time (e.g., Zwaan, Magliano, & Graesser, 1995). Recent theoretical work has also focused on accounting for processing along multiple dimensions. Zwaan and Radvansky (1998) propose a general framework for situation model processing in which they distinguish between the current model (i.e., the part of the model being constructed at time T_n) and the integrated model (i.e., the situation model that has been constructed from time T_1 through time T_{n-1}). A critical process in situation model construction involves incorporating the current model into the integrated model, which Zwaan and Radvansky refer to as *updating*. As a theoretical account of the updating process, they further describe the *event-indexing model* (Zwaan, Langston, & Graesser, 1995). A slightly more general formulation of the model can be stated as follows: consistent with previous research mentioned above, the events and entities represented in situation models can be related along several possible dimensions. Importantly, the ease of updating the integrated model – that is, the ease of incorporating the current model into the integrated model – will depend upon how many dimensions are shared between the entities and events of the current model and the integrated model. As predicted by the model, initial results show that rated coherence increases and reading times decrease additively with increases in the number of dimensions along which the content of a sentence is related to previous content.

The role of working memory in text comprehension

As should be apparent at this point, a great deal of processing is involved in text comprehension, including linguistic-level processes, semantic processes to form the text base, and knowledge retrieval and integration to form the situation model. An assumption common to all models of comprehension is that all information processing must take place in a finite capacity working memory. For instance, if two concepts never co-occur in working memory during the processing of a text, no new associations between these concepts will be formed as a consequence of reading this text. This postulate sets severe limits on the comprehension process, especially because the capacity of short-term memory, upon which working memory must rely, at least in part, is known to be quite small, only about four chunks.

If all processing depends on working memory, and working memory is so severely limited in terms of its capacity, variations of the capacity of working memory among individual readers ought to be closely related to comprehension. Daneman and Carpenter (1980) measured short-term memory capacity in the context of a reading task by asking subjects to read a series of sentences and then recall the last word from each sentence. Working memory measured in this way (referred to as *reading span*) correlates quite well with reading comprehension. Reading span differs among individuals, varying between about 2 and 6, and is a reliable predictor of performance on conventional reading comprehension tests (including the SAT), as well as inferencing (Singer, Halldorson, Lear, & Andrusiak, 1992). However, while the reading span is a good predictor of individual differences in reading comprehension, it can be argued that estimates of working memory capacity arrived at in this way are too low to be able to account for everything a good reader must maintain in working memory: crucial fragments of the prior text, including its macrostructure, linguistic knowledge, relevant world knowledge, reading goals – a list much too long for even the highest reading span yet encountered. An explanation for this apparent discrepancy was provided by Ericsson and Kintsch (1995) and Kintsch, Patel, and Ericsson (1999) who introduced the concept of long-term working memory. Working memory, when we are reading a text in a familiar domain, is composed not only of a limited-capacity short-term memory but also includes a long-term component. This component contains all items in the reader's long-term memory that are linked to the current contents of short-term memory (which for this purpose can be equated with consciousness or focus of attention) via retrieval structures. Thus, retrieval structures make available information stored in long-term memory that is directly relevant to the task at hand without the need for time- and resource-consuming retrieval processes.

Retrieval structures, however, are characteristic of performance in expert domains, and, indeed, are limited to expert domains. Long-term working memory allows the chess master to "see" the next move without having to figure it out; it makes it possible for the experienced physician to integrate patient data, medical knowledge, and prior experience to arrive at an intuitive diagnosis; and it is the basis for reading comprehension, for when we are reading texts in familiar domains, we are all experts, having practiced comprehension for many years. Retrieval structures exist only in domains where people have acquired expertise, which requires a great deal of practice. Thus, the chess master and the physician are not necessarily better than anyone else outside their domain of expertise. Most people reading a paper in theoretical physics do not readily comprehend it like they comprehend the daily newspaper, or as a physicist would comprehend the physics paper. Rather than fluent reading, their reading in the latter case would be an arduous and frustrating problem-solving activity.

The concept of long-term working memory allows us to understand how readers can juggle all the things they need for comprehension in working memory. It also suggests a reinterpretation of the reading span data: it is not the capacity of working memory that varies among individuals, but the skill with which they are using this capacity. High-span readers are readers who have a high level of reading expertise: they are fluent decoders who easily organize detailed information into a hierarchical macrostructure, and who possess rich, well-elaborated knowledge of word meanings. Their efficient processing allows them to effectively employ their retrieval structures as they comprehend and enables them to recall many words. In contrast, inefficient lower level reading processes is characteristic of readers with low reading spans.

Expert comprehension, thus, is a highly automated process, relying on readily available retrieval structures. To establish these retrieval structures, extensive practice is required. This raises a dilemma for the development of comprehension skills. Novice comprehension obviously cannot rely on automatized skills. Instead, active strategic processes must compensate for the lacking retrieval structures that make comprehension easy for the expert. The novice comprehender reading unfamiliar material must expend considerable effort, and expend it in just the right way, to achieve adequate results. A major difficulty for comprehension instruction therefore is to engage novice comprehension in the kind of active, strategic processing that is necessary to build good situation models. All too easily, novice comprehenders are satisfied with forming a reasonably accurate textbase, neglecting the more effortful construction of a situation model. But that results in shallow comprehension, which is insufficient for deep understanding and learning from texts.

Summary

Text comprehension is a complex process, requiring the involvement of many different components, relying upon many different kinds of information, and yielding complex mental representations. No one chapter can examine every process involved at a level of detail that would do justice to their importance and complexity. Instead, we hope to have met the more modest goals of highlighting the inherent complexity of text comprehension, introducing the reader to several of the key processes thought to underlie comprehension, and pointing to relevant bodies of literature for more interested readers. Indeed, many of the processes mentioned herein enjoy substantial bodies of research that have been directed at understanding how each component process operates. Almost all of the extant research on text comprehension has focused on identifying and examining the various component processes in isolation, which has gone a long way toward furthering our understanding of how comprehension works. However, text comprehension is not simply the sum of the activity of these various processes, but *arises from their coordinated operation as a system*. Thus, an important direction for future research will be to examine the interaction of the various components to understand how they work together as a system to give rise to comprehension.