Running head: DISCOURSE COMPREHENSION

Discourse Comprehension

Arthur C. Graesser and Carol Forsyth

University of Memphis

Graesser, A.C., & Forsyth, C. (in press). Discourse comprehension. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*. Oxford, UK: Oxford University Press.

Send correspondence to:

Art Graesser
Psychology Department
202 Psychology Building
University of Memphis
Memphis, TN, 38152-3230
901-678-4857
901-678-2579 (fax)
a-graesser@memphis.edu

KEYWORDS : Discourse processing; comprehension,

Abstract

Discourse comprehension is viewed from a multilevel framework that includes the levels of words, syntax, textbase, situation model, rhetorical structure, genre, and pragmatic communication. Discourse researchers investigate the cognitive representation of these levels and the process of constructing them during comprehension. Comprehension frequently is successful at all levels, but sometimes there are communication misalignments, ungrounded symbols, and other comprehension obstacles that increase processing time or that lead to comprehension breakdowns. Psychologogists have developed a number of psychological models of discourse comprehension, such as the the construction-integration, constructionist, and indexical-embodiment models. Advances in corpus and computational linguistics have allowed interdisciplinary researchers to systematically analyze the words, syntax, semantics, cohesion, situation model, world knowledge, and global structure of texts with computers. These computer analyses help researchers discover new discourse patterns, test hypotheses more rigorously, assess potential confounding variables, and scale texts on difficulty.

Discourse Comprehension

Our definition of discourse includes both oral conversation and printed text. The spoken utterances in oral conversation and the sentences in printed text are composed by the speaker/writer with the intention of communicating interesting and informative messages to the listener/reader. Therefore, naturalistic discourse is likely to be coherent, understandable to the community of discourse participants, and relevant to the situational goals. Sometimes discourse communication breaks down, however. Communication breakdowns occur when the writer and reader (or speaker and listener) are faced with substantial gulfs in language, prior knowledge, or discourse skills. Minor misalignments often grab the comprehender's attention, as in the case of a mispronounced word, a rare word in a text, an ungrammatical sentence, or a sentence that does not fit into the discourse flow. A model of discourse comprehension should handle instances when there are communication breakdowns in addition to successful comprehension.

Psychological theories of comprehension have identified the representations, structures, strategies, and processes at multiple levels of discourse (Clark, 1996; Graesser & McNamara, 2009; Graesser, Millis, & Zwaan, 1997; Kintsch, 1998; McNamara & Magliano, 2009; Pickering & Garrod, 2004; Snow, 2002; Van Dijk & Kintsch, 1983). The taxonomy we adopt in this chapter is an expanded version of one presented by Graesser et al. (1997): *words, syntax*, the explicit *textbase*, the referential *situation model* (sometimes called the mental model), the discourse *genre and rhetorical structure* (the type of discourse and its composition), and the *pragmatic communication* level (between speaker and listener, or writer and reader). Words and syntax are self explanatory and form what is sometimes called the surface code. The textbase contains explicit propositions in the text in a form that preserves the meaning but not the surface code of wording and syntax. The situation model is the referential content or microworld that the text is describing. This would include the people, objects, spatial setting, actions, events, processes, plans, thoughts and emotions of people, and other referential content. The text genre is the type of discourse, such as a news story, a folk tale, a persuasive editorial, or a science text that explains a causal mechanism.

Table 1 elaborates on these six levels by identifying the codes, constituents, and content associated with each level. This chapter will not crisply define each level and the associated terminology, but the table does provide examples of what each level contains. It is sometimes debatable which level to assign a particular component of language or discourse; this is quite expected because of interactions between levels. Moreover, Table 1 depicts the levels of discourse as compositional components that are constructed as a *result* of comprehension. It is important not to lose sight of the fact that this *compositional* viewpoint is incomplete without considering the affiliated *knowledge* and *process* viewpoints. For any given compositional entity C, the person needs to have had the prerequisite knowledge about C through prior experiences and training. The person needs to be able to process C by identifying its occurrence in the discourse and by executing relevant cognitive processes, procedures, and strategies proficiently.

INSERT TABLE 1 ABOUT HERE

The remainder of this chapter has three sections. The first section discusses some of the mechanisms that operate when readers/listeners experience comprehension difficulties, breakdowns, or communication misalignments. Such comprehension challenges are informative because they predict measures of attention, reading time, memory, reasoning, behavior, and other manifestations of cognition. The second section reviews psychological models of discourse that attempt to explain comprehension processes and representations. The third section describes advances in corpus and computational linguistics that have computers automatically analyze texts on the various discourse levels: words, syntax, semantics, cohesion, situation models, global structure, and world knowledge. These three sections are written from the lens of the multilevel discourse comprehension framework.

Obstacles of Multilevel Comprehension

Comprehenders can face obstacles at any of the levels in Table 1. There can be deficits in the comprehender (i.e., lack of knowledge or skill) or the discourse (e.g., incoherent text, unintelligible speech). The severity of the obstacle can range from a minor irregularity that adds some cost in processing time to a complete breakdown in comprehension. Attempts can be made to compensate for the problem by recruiting information from other levels of discourse, from prior knowledge, from external sources (e.g., other people or technologies), or from strategies. Sometimes deeper levels of comprehension can compensate for deficits at the shallower levels, but that will not help if the words and shallow levels carry the burden of information constraints. The scenarios below illustrate some discourse obstacles and resulting consequences.

Scenario 1. A student in a foreign language course has mastered the phonemes but very little of the vocabulary. The vocabulary deficit prevents him from understanding any of the conversations in class. The breakdown at discourse level 1 also blocks the deeper levels of 2-6 (see Table 1).

Scenario 2. An employee reads a health insurance document that has lengthy sentences with embedded clauses and numerous quantifiers (*all, many, rarely*) and Boolean operators (*and, or*, *not, if*). She understands nearly all of the words, but has only a vague idea what the document explicitly states because of complex syntax, a dense textbase, and an ungrounded situation model (i.e., deficits at levels 2-4). However, she signs the contract because she understands its purpose and trusts the Human Resources Department of the employer. Levels 5 and 6 circumvent the need to understand levels 2-4 completely.

Scenario 3. A couple read the directions to assemble a new computer. They argue about how to hook up the cables on the dual monitors. They have no problem understanding the words and textbase in the directions (levels 1-3) and no problem understanding the genre and purpose of the document (levels 5 and 6), but they do have a deficit at the situation model level (level 4).

Scenario 4. A science student asks his roommate to proofread a term paper, but the roommate is a journalism major who knows little about science and complains that there is a problem with logical flow. The science major revises the text by adding connectives (e.g., *because, so, therefore, before*) and other words to improve the cohesion. The revised composition is deemed more comprehensible. In this case, improvements at levels 1 and 3 compensated for a deficit at level 4.

. **Scenario 5**. Parents take their children to a new Disney movie that they discover has a few adult themes. The children notice the parents laughing at different points in the movie than they do. The children are making it successfully through discourse levels 1-4, but levels 5 and 6 are not intact.

These scenarios illustrate how deficiencies at one or more discourse levels can have substantial repercussions on the processing at other levels. Discourse researchers need to understand the processing mechanisms both *within* levels and *between* levels.

Metacognition is the knowledge that person has about cognitive mechanisms – thinking about thinking (Hacker, Dunlosky, & Graesser, 2009). Research on meta-comprehension has revealed that people may have deficits at one or more comprehension levels without being aware of it. For example, in research on *comprehension calibration* (Dunlosky & Lipko, 2007; Maki, 1998), ratings are collected from readers on how well they believe they have comprehended texts and these ratings are correlated with objective tests of text comprehension. The comprehension calibration correlations are alarming low (r = .27) even among college students. Readers often have an illusion of comprehension when they read text because they settle for shallow levels of analysis as a criterion for adequate comprehension (Baker, 1985; Daneman, Lennertz, & Hannon, 2006; Otero & Kintsch, 1992). Shallow readers believe they have adequately comprehended text if they can recognize the content words and can understand most of the sentences, when in fact they are missing the deeper knowledge and occasional contradictions or false claims. Deep comprehension requires inferences, linking ideas coherently, scrutinizing the validity of claims with a critical stance, and understanding the motives of authors (Kendeou & Van den Broek, 2007; Rapp, 2008; Rouet, 2006; Wiley et al., 2009). Deep comprehension may only be selectively achieved in everyday comprehension experiences. Many readers settle for shallow comprehension for discourse unless they have a high amount of background knowledge (O'Reilly & McNamara, 2007a), the information is in the discourse focus (Sanford & Graesser, 2006; Ward & Sturt, 2007), the genre dictates careful scrutiny of information (Kendeou & van den Broek, 2007), or the information is highly relevant to the readers' goals (Kaakinen & Hyona, 2007).

In studies conducted in our laboratory (Graesser et al., 2004; VanLehn et al., 2007), college students read textbooks on technical topics such as computer literacy and Newtonian physics. They subsequently completed a rigorous test on deep knowledge with multiple-choice questions similar to the Force Concept Inventory in physics (Hestenes, Wells, & Swackhamer, 1992). We were surprised to learn that the college students had zero learning gains from reading the textbook and that the posttest scores did not differ from a condition in which the students read nothing at all. In contrast, the learning gains were quite substantial when there was a learning environment that challenged their comprehension of the material and engaged in tutorial dialogue (through a computer system called AutoTutor). Results such as these strongly suggest that the reading strategies of literate adults are far from optimal when considering deep comprehension. Our college students did not achieve deep comprehension on texts about physics and computer literacy even when they had a nontrivial amount of world knowledge on these topics and sufficient reading strategies to land them in college.

Comprehension deficits are prevalent in conversations. An important foundation for communication is forming a *common ground* (shared knowledge) among speech participants (Bard et al., 2007; Clark, 1996; Holler & Wilkin, 2009; Schober & Brennan, 2003). This requires making appropriate assumptions on what each other already knows ("old" or "given" knowledge in the common ground) and by performing acts of communication (verbal or nonverbal) about new information that is properly *coordinated* with what each other knows. People normally assume that knowledge is in the common ground if it (a) is physically co-present and salient (all parties in the communication can perceive it), (b) has been verbally expressed and understood in the discourse space, in front of the sender, recipients, and side audience, and/or (c) it is common knowledge for members of a group, culture, or target community. Communication breakdowns occur to the extent that one or more of these conditions are in jeopardy.

Regarding the coordination of acts of communication, there are discourse devices that facilitate successful communication. Clark (1996) has proposed four levels in the *joint action ladder*:

Level A: Attention. Is the intended recipient paying attention?

Level B: Listening. Is the recipient actively listening and identifying signals?

Level C: Understanding. Does the recipient understand the message of the sender?

Level D: Action. Does the recipient perform a verbal or physical action that reflects understanding?

Breakdowns occur when there is a misfire at any one of these levels. However, it can be difficult to know which level is problematic when there is no response from a recipient. Clark's *principle of upward causality* predicts that disruption at lower levels propagate to higher levels. Indeed, failure at the attention Level A accounts for many of the disruptions that occur in communication technologies, such as email and instant messaging (Hancock & Dunham, 2001). Simply put, it's not clear whether the recipient is paying attention. The sender can ask questions to inquire about the status of the listener with respect to these four levels: Are you there? Are you listening? Do you understand? Could you recap/summarize? These discourse acts correspond to a second track discussion (Track 2, meta-communication) that specifically addresses communication problems.

One important signal of a recipient is *backchannel feedback*. Backchannel feedback ("uh huh," "okay," head nod) from the recipient acknowledges the sender's message (i.e., conversation Levels 1 and 2 are in check). In face-to-face conversations in English, there should be back channel feedback from the recipient after every 10-15 syllables, on the average. The recipient who gives more backchannel feedback can be viewed as annoying whereas much less feedback makes the recipient appear distracted, skeptical, or unresponsive. In email, backchannel feedback is an important courtesy after a sender's message, but some recipients do not extend this courtesy. Sometimes it is appropriate to withhold backchannel feedback, as in the case of a large teleconference on a telephone system that imposes transmission latencies after every party turn; when this occurs the organizer needs to declare rules such as “Whenever I say something and then pause, please respond only if you disagree or have something new to say.”

Comprehension is manifested in *conversation patterns* with larger sequences of turns between conversational participants. For example, in the context of classroom teaching, there is the IRE sequence (Sinclair & Coulthard, 1975) in which the instructor asks a question (Initiate), the student answers (Response), and the instructor gives feedback on the answer (Evaluate). In tutorial dialogue, there is a 5-step tutoring frame in which the tutor asks a difficult question, the student answers, the tutor gives short feedback (positive, negative or neutral), there is an optional extended multi-turn elaboration of the answer, the tutor asks a comprehension-gauging question (“Do you understand?”), and the student designates their level of understanding (Graesser, D’Mello, & Person, 2009). In dialogues between a leader and follower that require referential grounding, there often are 4-step sequences: Leader: “You lift the red bar.” Follower: “The red bar?” Leader: “Yes, that red bar.” Follower: “Okay, the red bar” (Clark & Wilkes-Gibbs, 1986). This 4-step frame can be more economically expressed in face-to-face interaction by pointing gestures, directing a person’s gaze, and other non-verbal channels (Bard et al., 2007; Hanna & Brennan, 2007; Holler & Wilkin, 2009; Van der Sluis & Krahmer, 2007).

Available psychological research supports a number of generalizations about the processing order, constraints, interaction, and compensatory mechanisms of the different levels of discourse comprehension. Some of these generalizations are briefly described below.

(a) **Bottom-up dependencies of meaning**. In a strictly bottom-up model of reading, the ordering on depth is assumed to be levels 1 → 2 → 3 → 4 → 5 → 6 (see Table 1). Most researchers endorse an interactive model of reading rather than a strictly bottom-up model (Rayner & Pollatsek, 1994; Rumelhart, 1977; Taraban & McClelland, 1988; van den Broek, Rapp, & Kendeou, 2005). However, they also assume asymmetry in the constraints, such that the lower levels constrain the higher levels more than vice versa. The reading of words is robustly influenced by the bottom-up constraints of the letters and syllables (Gough, 1972; Rayner, 1998; Rayner & Pollatsek, 1994). The quality of a person's lexicon has a large impact on the proficiency and speed of interpreting sentences and generating inferences at levels 2 and higher (Perfetti, 2007; Stanovich, 1986). It is important for readers to establish an interpretation of the textbase before they can productively move on to the construction of the situation model and higher levels (O'Reilly & McNamara, 2007b). A partial-to-full analysis of levels 1-4 is presumably needed to adequately construct the rhetorical structure.

There is some question about the extent to which top-down processes influence lower level processes. Top-down processing is known to influence the speed and construction of word meanings (Rayner & Pollatsek, 1994), but there is less certainty about top-down influences on the construction of the textbase and situation model. Bransford and Johnson (1972) conducted a series of studies with vague abstract texts that were very difficult to comprehend without the introduction of a title that identifies a higher order schema (such as washing clothes) that coherently organizes the text content. Zwaan (1994) reported that the encoding of the surface code, textbase, and situation model had different profiles when college students were told they were reading a newspaper article versus literature. As predicted, the literature instructions enhanced the surface code whereas the newspaper instructions enhanced the situation model. These top-down influences of comprehending the textbase and situation model are provocative, but the effects are confined to texts that are extremely ambiguous or malleable in interpretation. The vast majority of texts are much more constrained.

(b) **Novel information requires more processing effort than familiar and automatized components**. Novelty of information is a foundational cognitive dimension that attracts attention, effort, and processing resources, and that predicts salient in memory (Tulving & Kroll, 1995). In spoken conversation, there are prosodic features that signal the occurrence of novel information and new information in the discourse space (Clark, 1996; Nygaard, Herald, & Namy, 2009; Riesco-Bernier & Romero-Trillo, 2008). Reading studies with eye tracking or self-paced reading times show that more processing time is allocated to rare words than high frequency words (Just & Carpenter, 1987; Pollatsek, Slattery, & Juhasz, 2008; Rayner, 1998) and to new information expressed in the textbase and situation model than to old information already mentioned (Haberlandt & Graesser, 1985; Kaakinen & Hyona, 2007). Graesser and McNamara (2009) proposed that the highest density of novel information resides in lower frequency words, the textbase level, and the situation model (levels 1, 3, and 4). In contrast, most aspects of syntax, genre, and author characteristics (levels 2, 5, and 6) are frequently experienced and therefore overlearned and automatized; these levels are quickly processed or are invisible to the comprehender unless there are obstacles.

(c) **Attention, consciousness, and effort gravitate to comprehension obstacles**. Obstacles at any level of analysis are likely to draw cognitive resources. Reading time studies have shown that extra processing time is allocated to pronouns that have unresolved or ambiguous referents (Gernsbacher, 1990; Rayner, 1998), to sentences that have breaks in textbase cohesion (Gernsbacher, 1990), to sentences that have coherence breaks in the situation model on the dimensions of temporality, spatiality, causality, and intentionality (Magliano & Radvansky, 2001; Zwaan, Magliano & Graesser, 1995; Zwaan & Radvansky, 1998), and to sentences that contradict ideas already established in the evolving situation model (Kendeou &

van den Broek, 2007; O'Brien, Rizzella, Albrecht, & Halleran, 1998; Rapp, 2008). Attention drifts toward sources of cognitive disequilibrium, such as impasses, anomalies, discrepancies, and contradictions (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005).

(d) **Comprehension obstacles may be repaired or circumvented by world knowledge, information at other discourse levels, or external sources**. The scenarios described earlier illustrate some compensatory mechanisms that repair or circumvent the comprehension obstacles. For example, the syntax, textbase, and situation model deficits in scenario 2 are circumvented by the information in discourse levels 5 and 6; in this case the person has enough information about levels 5 and 6 to know that deep understanding of levels 2-4 are unnecessary. The gaps and misalignments in the situation model of scenario 3 are rectified by extended conversations between the couple and by active problem solving. Coherence gaps in the textbase and situation model of scenario 4 are rectified by augmenting the discourse at levels 1 and 2 with connectives and other cohesion markers. Inserting these connectives and markers are known to improve comprehension, particularly for readers with low subject matter knowledge or low reading comprehension skill (Britton & Gulgoz, 1991; McNamara & Kintsch, 1996; O'Reilly & McNamara, 2007b).

The multilevel comprehension framework outlined in this section has provided a plausible sketch of the complexities of constructing meaning on different levels during discourse comprehension. There are multiple levels of meaning that mutually, but asymmetrically, constrain each other. The components at each level are successfully built if the text is naturalistic and the reader has prerequisite background knowledge and reading skills. However, there are periodic comprehension obstacles that range from minor misalignments and comprehension difficulties to complete communication breakdowns. The misfires are magnets of attention that sometimes trigger compensatory mechanisms that repair or circumvent the problems.

Psychological Models of Discourse Comprehension

Discourse psychologists have developed several theoretical models of comprehension during the last two decades. It is beyond the scope of this chapter to cover all of these models, but we will contrast three models that are representative of particular classes of models. A construction-integration model (Kintsch, 1998) will represent a class of bottom-up models, which would also include the memory-based resonance model developed by Myers and O'Brien (Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien et al., 1998). A constructionist model by Graesser, Singer, and Trabasso (1994) will represent a class of strategy-driven models, which would also include the structure-building framework (Gernsbacher, 1990) and the event indexing model (Zwaan et al., 2005; Zwaan & Radvansky, 1998). An indexical model by Glenberg (Glenberg & Robertson, 1999) will represent a class of embodied cognition models (Glenberg, 1997; de Vega, Glenberg, & Graesser, 2009). There are other models that that can be viewed as hybrids of these three classes, such as the landscape model (Van den Broek, Virtue, Everson, Tzeng, & Sung, 2002), the CAPS/Reader model (Just & Carpenter's, 1992), and the 3CAPS model (Goldman, Varma, & Cote, 1996).

Construction-Integration Model

Kintsch's (1998) construction-integration (CI) model is currently regarded as the most comprehensive psychological model of comprehension. The model accommodates a large body of psychological data, including reading times, activation of concepts at different phases of comprehension, sentence recognition, text recall, and text summarization. Comprehension strategies exist, but they do not drive the comprehension engine. Instead, comprehension lies in

(a) the bottom-up activation of knowledge in long-term memory from textual input (the *construction* phase) and (b) the integration of activated ideas in working memory (the *integration* phase). As each sentence or clause in a text is comprehended, there is a construction phase followed by an integration phase.

The construction phase for each sentence activates hundreds of *nodes,* which correspond to concepts, propositions, rules, and other forms of content. The nodes cover the various levels of representation in the multilevel comprehension framework (see Table 1). The model assumes that a connectionist network (Mayberry, Crocker, & Keller, 2009; Taraban & McClelland, 1988) is iteratively created, modified, and updated during the course of comprehension. That is, as text is read, sentence by sentence (or clause by clause), a set of word concept nodes and proposition nodes are activated (constructed). Some nodes correspond to explicit constituents in the text whereas others are activated inferentially. The activation of each node in the network fluctuates systematically during the course of comprehension as each sentence is read. When a sentence (or clause) S is read, the set of N activated nodes include (a) the explicit and inference nodes affiliated with S and (b) the nodes that are held over in working memory from the previous sentence S-1 by virtue of meeting some threshold of activation. There are N nodes that have varying degrees of activation while comprehending sentence S. These N nodes are fully connected to each other in a weight space. The set of weights in the resulting N by N *connectivity matrix* specifies the extent to which each node activates or inhibits the activation of each of the N nodes. The values of the weights in the connectivity matrix are theoretically motivated by the multiple levels of language and discourse. For example, if two proposition nodes (A and B) are closely related semantically, they would have a high positive weight, whereas if the two propositions contradict each other, they would have a high negative weight.

The integration phase modifies activation of the N nodes dynamically. At construction, the N nodes are activated to varying degrees, specified by an initial activation vector (a_1, a_2, …a_N). The connectivity matrix then operates on this initial node activation vector in multiple activation cycles until there is a settling of the node activations to a new final stable activation profile for the N nodes. At that point, integration of the nodes has been achieved. This is computed mathematically by the initial activation vector being multiplied by the same connectivity matrix in multiple iterations until the N output vectors of two successive interactions shows extremely small differences (signifying a stable settling of the integration phase). Sentences that are more difficult to comprehend would presumably require more cycles to settle. The settling process history and/or final activation values of the N notes are able to predict different types of experimental measures, such as reading times, word priming, recall, recognition, and summarization.

Constructionist Model

Comprehension is more directed and strategic according to the constructionist theory proposed by Graesser, Singer, and Trabasso (1994). The distinctive strategies of this model are reflected in its three principal assumptions: reader goals, coherence, and the explanation. The *reader goal* assumption states that readers attend to content in the text that is relevant to the goals of the reader. Newspapers are read for very different purposes, and never completely, front to back. The *coherence* assumption states that readers attempt to construct meaning representations that are coherent at both local and global levels. Therefore, coherence gaps in the text will stimulate the reader to actively think, generate inferences, and reinterpret the text in an effort to fill in, repair, or take note of the coherence gap. The *explanation* assumption states that good comprehenders tend to generate explanations of *why* events and actions in the text occur, *why* states exist, and *why* the author bothers expressing particular ideas. Why-questions encourage analysis of causal mechanisms, justifications of claims, and other deeper levels of the situation model. There are other assumptions of the constructionist theory that are shared by many other models, assumptions that address memory stores, levels of representation, world knowledge, activation of nodes, automaticity, and so on, but its signature assumptions address reader goals, coherence, and explanation.

As in the case of the construction-integration model, the constructionist theory has accounted for data involving reading times, word priming, inference generation, recall of text information, and summarization. The notion that coherence and explanation strategies are the hallmarks of good comprehension places constraints on comprehension. These strategies determine the selection of content that gets encoded, the inferences that are generated, and the time spent processing text constituents. For example, proficient readers are driven by why-questions more than how, when, where, and what-if, unless there are special goals to track such information. The explanations of the motives of characters and of the causes of unexpected events in a story are much more important than the spatial position of the characters in a setting, what the character looks like, and the procedures and style of how characters' actions are performed. Such details about space, perceptual attributes, and actions are important when they serve an explanatory function or they address specific reader goals. When readers are asked to monitor why-questions during comprehension, their processing and memory for the text is very similar to normal comprehension without such an orienting questions; however, when asked to monitor how questions and what happens next questions, their processing and memory shows signs of being disrupted (Magliano, Trabasso, & Graesser, 1999). Explanations and why-questions are fundamental to the construction of meaning according to the constructionist model.

Research has confirmed that students comprehend text deeper when they generate self-explanations of the materials (Chi, de Leeuw, Chiu, & LaVancher, 1994; McNamara, 2004; Millis et al., 2004).

Indexical Hypothesis and Embodiment

Glenberg's *Indexical Hypothesis* (Glenberg & Robertson, 1999) adopts an embodied theory of language and discourse comprehension (Glenberg, 1997; Glenberg & Kaschak, 2004; Pecher & Zwaan, 2005). The central theoretical claim is that meaning is grounded in how we use our bodies as we perceive and act in the world. For example, comprehension of a story is predicted to improve after children have been able to perceive and manipulate the characters and objects in a story scenario. When adults read a manual on assembling a piece of equipment, their comprehension is expected to improve to the extent that they can enact the procedures or at least form visual images of the objects and actions. Readers who have the metacognitive strategy of grounding the entities and events mentioned in the text are expected to show comprehension advantages over those who do not bother taking such extra cognitive steps. It should be noted that the constructionist model would not encourage these strategies unless they served the strategies of building explanations, coherent representations, and representations that address particular reader goals. Similarly, the construction-integration model would not directly predict the importance of embodied representations.

A recent edited volume published some debates on the conceptual differences and the empirical evidence for embodied versus symbolic theories of comprehension (deVega, Glenberg, & Graesser, 2008). A *strong* sense of embodiment exists in a representation that incorporates the constraints of an organism's body, its location in the world, its perspective in perceiving the world, and its perceptual-motor interactions with the world. A *weak* sense of embodiment exists when there are vestiges of perceptions, actions, and perspectives in the representation, but the components are less detailed or underspecified, yet to some extent systematic or recoverable. A representation is *not* embodied when the symbols have an arbitrary relationship with the various components of perceptual-motor interactions with the world. In contrast, a symbolic representation is a structured set of symbols, each of which stands for some aspect of a referential domain. What it stands for may or may not be embodied. An *amodal* symbolic representation is not grounded in any embodied representation. A *modal* symbolic representation is connected to perceptual-motor experience either indirectly through interpretive mechanisms or directly through sensory transduction and motoric actuators. Consequently, embodied and symbolic representations are not necessarily mutually exclusive.

There is growing support for the embodied framework even though it has enjoyed only a decade of empirical testing (Pecher & Zwaan, 2005; Masson, Bub, & Warren, 2008; de Vega et al., 2008). However, as would be expected, there are some fundamental challenges for the embodiment framework in explaining discourse comprehension. The first challenge is that it is difficult to explain how embodied representations can be constructed at a normal reading rate of 150-400 words per minute. It takes approximately 300-1000 milliseconds to construct a new referent in a discourse space (i.e., in the mind's eye), several hundred milliseconds to move an entity from one location to another, a few hundred more milliseconds to have the mental camera zoom in on an entity within a crowded mental space, and so on (Kosslyn, 1980; Millis, King, & Kim, 2001). These considerations on timing and complexity raise some doubts that all referring expressions and clauses in the text have fully embodied representations during reading. It should be noted that this challenge about comprehension time also would apply to the constructionist theorists who claim that deep explanations are constructed during comprehension. It may be impossible to construct deep explanations at a reading rate of 150-400 words per minute, which would explain the results of the studies that were reported earlier that very little deep knowledge is acquired from reading textbooks.

The second challenge is that embodied representations are constructed only under very restricted conditions. Graesser and Jackson (2008) have argued that the embodied framework is essentially correct under the following conditions: (1) when tasks and tests encourage embodied activities, (2) when the stimuli are simple (e.g., few actors and objects in the mind's eye), (3) when there is an existing visual-spatial grounding (e.g., an established spatial layout), and (4) when there is sufficient time and cognitive resources to carry out these processing operations. It is plausible that the small amount of content in the discourse focus is an excellent candidate for being a recipient of such cognitive activities. In contrast, disembodied symbolic representations are more explanatory when the relevant task goals do not encourage embodied processing, when the stimulus is complex, when there is minimal visual-spatial grounding, and when the reading rate is at the fast end of 150-400 words per minute. Much of the content that is presupposed and highly embedded will not be a good candidate for becoming a fully fleshed out embodied representation.

The above analysis addresses the relatively time-consuming integration phase of Kintsch's construction integration model (Kintsch, 1998) rather than the initial activation of representations. It is possible to have quick activations of many types of representations, both embodied and symbolic, during the initial activation of information associated with content words. Much of this automatic activation of representations end up dying away and never make it to the integration phase that establishes a more coherent representation of the meaning of the text.

Computer Tools for Analyzing Language and Discourse at Multiple Levels

This is a unique point in history because there is widespread access to hundreds of computer tools that analyze specific texts and large text corpora. This increase in automated text analyses can be attributed to landmark advances in computational linguistics (Jurafsky & Martin, 2008), discourse processes (Graesser, Gernsbacher, & Goldman, 2003), statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007), and corpus analyses (Biber, Conrad, & Reppen, 1998). Thousands of texts can be quickly accessed and analyzed on thousands of measures in a short amount of time. Of course, many theoretical components of discourse cannot be automated. In such cases it is necessary to have human experts annotate the texts systematically. However, human annotation is an expensive and time-consuming alternative, so it is essential to off load much of the work to computers. Moreover, an objective analysis of discourse should not rely entirely on human intuitions for scoring and annotation.

This chapter will present recent work on automated text analysis through the lens of Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, McCarthy, & Graesser, in press). Coh-Metrix is a computer facility that analyzes texts on most of the discourse levels in Table 1. More specifically, Coh-Metrix was developed to analyze and measure text on levels 1 through 5. The original purpose of the Coh-Metrix project was to concentrate on the cohesion of the textbase and situation model because those levels needed a more precise specification. However, it quickly became apparent that there is a need to automatically measure language and discourse processing at all of the levels under the rubric of the multilevel comprehension framework. The theoretical vision behind Coh-Metrix was to use the tool to (a) assess the overall cohesion and language difficulty of discourse on multiple levels, (b) investigate the constraints of discourse within levels and between levels, and (c) test models of multilevel discourse comprehension. There were also some practical goals in our vision: (a) to enhance standard text difficulty measures by providing scores on various cohesion and language characteristics and (b) to determine the appropriateness of a text for a reader with a particular profile of cognitive characteristics.

Coh-Metrix is available in both a public version for free on the web (http://cohmetrix.memphis.edu, version 2.0) and an internal version (versions 2.1 and 3.0). The public version has over 60 measures of language and discourse at levels 1-5 in Table 1, whereas the internal research version has nearly a thousand measures that are at various stages of testing. Coh-Metrix is used by simply entering a text, filling in identifier information about the text, and clicking on a button. After a few seconds, the system produces a long list of measures on the text. If the text is extremely lengthy, the text can be divided into textiles of 500-1000 words. There is a help system that defines the measures and that provides various forms of contextual support. Discussed below are measures associated with the various levels of the multilevel framework.

INSERT TABLE 2 ABOUT HERE

Words

Coh-Metrix was designed to move beyond standard readability formulas that rely on word length and sentence length to difficulty. Widely adopted measures of text difficulty are the Flesch-Kincaid Grade Level (Klare, 1974-5), Degrees of Reading Power (DRP; Koslin, Zeno, & Koslin, 1987), and Lexile scores (Stenner, 1996). Formula 1 shows the Flesch-Kincaid Grade Level metric. *Words* refers to mean number of words per sentence and *syllables* refers to mean number of syllables per word.

$$\text{Grade Level} = .39 * \text{Words} + 11.8 * \text{Syllables} - 15.59 \quad (1)$$

The lengths of words and sentences no doubt have important repercussions on psychological processes but we need Coh-Metrix to scale texts on more levels.

Many Coh-Metrix measures refer to characteristics of individual words. Much can be discovered from computer facilities that link words to psychological dimensions, as in the case of *WordNet* (Fellbaum, 1998) and *Linguistic Inquiry Word Count* (Pennebaker, Booth, & Francis, 2007). Coh-Metrix measures words on dozens of characteristics that were extracted from established psycholinguistic and corpus analyses, including WordNet and many of the categories of LIWC. The MRC Psycholinguistic Database (Coltheart, 1981) is a collection of human ratings of several thousands of words along several psychological dimensions: meaningfulness, concreteness, imagability, age of acquisition, and familiarity. Coh-Metrix computes scores for word frequency, ambiguity, abstractness, and parts of speech, as is documented in the Coh-Metrix help system. There is a relative frequency per 1000 words for each particular category of words.

Syntax

Coh-Metrix analyzes sentence syntax with the assistance of a syntactic parser developed by Charniak (2000). The parser assigns part-of-speech categories to words and syntactic tree structures to sentences. There are two notable measures of syntactic complexity that are predicted to place a high load on working memory. First, the *number of modifiers per noun-phrase* is an index of the complexity of referencing expressions. “The very large angry dog” is a noun-phrase with 4 modifiers of the head noun “dog.” Second, the number of words before the main verb of the main clause is an index of syntactic complexity because it places a burden on the working memory of the comprehender (Graesser, Cai, Louwerse, & Daniel, 2006; Just & Carpenter, 1987, 1992). Sentences with pre-posed clauses and left-embedded syntax require comprehenders to keep many words in working memory before getting to the meaning of the main clause.

Textbase

The *textbase* theoretically contains explicit propositions in the text, referential links between explicit propositions, and a small number of inferences that connect the explicit propositions (van Dijk & Kintsch, 1983). The propositions are in a stripped down form that removes surface code features captured by determiners, quantifiers, tense, aspect and auxiliary verbs. *Co-reference* is an important linguistic method of connecting propositions, clauses, and sentences in the textbase (Britton & Gulgoz, 1991; Halliday & Hasan, 1976; McNamara & Kintsch, 1996; van Dijk & Kintsch, 1983). Referential cohesion occurs when a noun, pronoun, or noun-phrase refers to another constituent in the text. For example, in the sentence *When the intestines absorb the nutrients, the absorption is facilitated by some forms of bacteria,* the word *absorption* in the second clause refers to the event (or alternatively the verb *absorb*) in the first clause. There is a referential cohesion gap when the words in a sentence or clause do not connect to other sentences in the text.

Coh-Metrix tracks five major types of lexical co-reference: *common noun overlap*, *pronoun overlap*, *argument overlap*, *stem overlap*, and *content word overlap*. Common noun overlap is the proportion of all sentence pairs that share one or more common nouns, whereas pronoun overlap is the proportion of sentence pairs that share one or more pronoun. Argument overlap is the proportion of all sentence pairs that share common nouns or pronouns (e.g., *table/table, he/he,* or *table/tables*). Stem overlap is the proportion of sentence pairs in which a noun (or pronoun) in one sentence has the same semantic morpheme (called a lemma) in common with any word in any grammatical category in the other sentence (e.g., the noun *photograph* and the verb *photographed*). The fifth co-reference index, content word overlap, is the proportion of content words that are the same between pairs of sentences. There are different variants of the five measures co-reference. Some indices consider only pairs of *adjacent* sentences, whereas others consider *all possible pairs* of sentences in a paragraph.

Coh-Metrix treats pronouns carefully because pronouns are known to create problems in comprehension when readers have trouble linking the pronouns to referents. Coh-Metrix computes the incidence scores for personal pronouns (I, you, we) and the proportion of noun-phrases that are filled with any pronoun (including it, these, that). *Anaphors* are pronouns that refer to previous nouns and constituents in the text. There are measures of anaphor overlap in Coh-Metrix that approximate binding the correct referent to a pronoun, but the pronoun resolution mechanism is not perfect. A pronoun is scored as having been filled with a referent corresponding to a previous constituent if there is any prior noun that agrees with the pronoun in number and gender and that satisfies some syntactic constraints (Lappin & Leass, 1994).

Connectives and discourse markers that have the special function of linking clauses and sentences in the textbase (Halliday & Hasan, 1976; Louwerse, 2001; Sanders & Noordman, 2000). The categories of connectives in Coh-Metrix include additive (*also, moreover*), temporal (*and then, after, during*), causal *(because, so*), and logical operators (*therefore, if, and, or*). A higher incidence of these connectives increase cohesion in the textbase and also the situation model.

Situation model

Text comprehension researchers have investigated five dimensions of the situational model (Zwaan et al., 1995; Zwaan & Radvansky, 1998): causation, intentionality, time, space, and protagonists. A break in cohesion or coherence occurs when there is a discontinuity on one or more of these situation model dimensions. Such discontinuities are known to increase reading times and trigger the generation of elaborative inferences (Zwaan & Radvansky, 1998). Whenever such discontinuities occur, it is important to have connectives, transitional phrases, adverbs, or other signaling devices that convey to the readers that there is a discontinuity; we refer to these different forms of signaling as *particles*. Cohesion is facilitated by particles that clarify and stitch together the actions, goals, events, and states in the text.

Coh-Metrix analyzes the situation model dimension on causation, intentionality, space, and time, but not protagonists. For *causal and intentional cohesion*, Coh-Metrix computes the ratio of cohesion particles to the incidence of relevant referential content (i.e., main verbs that signal state changes, events, actions, and processes, as opposed to states). The ratio metric is essentially a conditionalized incidence of cohesion particles: Given the occurrence of relevant content (such as clauses with events or actions, but not states), Coh-Metrix computes the density of particles that stitch together the clauses. For example, the referential content for intentional information includes intentional actions performed by agents (*kill, help, give*, as in stories, scripts, and common procedures); in contrast, the intentional cohesion particles include infinitives and intentional connectives (*in order to, so that, by means of*).

Measuring *temporal cohesion* is important because of its ubiquitous presence in organizing language and discourse. Time is represented through inflected tense morphemes (e.g., -*ed, is, has*) in sentences of the English language. The temporal dimension also depicts unique internal event timeframes, such as an event that is complete (i.e., *telic*) or ongoing (i.e., *atelic*), by incorporating a diverse tense-aspect system. The occurrence of events at a point in time can be established by a large repertoire of adverbial cues, such as *before, after, then*. The temporal measures of Coh-Metrix compute a repetition score that tracks the consistency of tense (e.g., *past* and *present*) and aspect (*perfective* and *progressive*) across a passage of text. The repetition scores decrease as shifts in tense and aspect are encountered. A low score indicates that the representation of time in the text is disjointed, thus having a possible negative influence on the construction of a mental representation. When such temporal shifts occur, the readers would encounter difficulties without explicit particles that signal such shifts in time, such as the temporal adverbial (*later on*), temporal connective (*before*), or prepositional phrases with temporal nouns (*on the previous day*). A low particle-to-shift ratio is a symptom of problematic temporal cohesion.

In addition to the co-reference variables discussed earlier, Coh-Metrix assesses conceptual overlap between sentences by a statistical model of word meaning: Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007). LSA is an important method of computing the conceptual similarity between words, sentences, paragraphs or texts because it considers implicit knowledge. LSA is a mathematical, statistical technique for representing world knowledge, based on a large corpus of texts. The central intuition is that the meaning of a word is captured by the company of other words that surround it in naturalistic documents. Two words are similar in meaning to the extent that they share similar surrounding words. For example, the word *glass* will be highly associated with words of the same functional context, such as *cup*, *liquid*, and *pour*. LSA uses a statistical technique called singular value decomposition to condense a very large corpus of texts to 100-500 statistical dimensions (Landauer et al., 2007). The conceptual similarity between any two text excerpts (e.g., word, clause, sentence, text) is computed as the geometric cosine between the values and weighted dimensions of the two text excerpts. The value of the cosine typically varies from 0 to 1. LSA-based cohesion was measured in several ways in Coh-Metrix, such as LSA similarity between adjacent sentences, LSA similarity between all possible pairs of sentences in a paragraph, and LSA similarity between adjacent paragraphs. The statistical representation of words in LSA depends on the corpus of texts on which they are trained. Coh-Metrix has different corpus options, but the default is the TASA corpora of academic textbooks; this is a corpus of over 10 million words that cover a broad range of topics.

Coh-Metrix supplies a LSA-based measure of *given versus new information* in text. Each sentence has a measure of the amount of given (G) versus new (N) information and a proportion score is computed for newness [N/(N+G)]. A text with a low newness score is considered redundant.

Genre and Rhetorical Composition

Coh-Metrix distinguishes texts in three genres that are typical of high-school reading exercises: *narrative*, *social studies*, and *science*. The indices are derived from discriminant analyses conducted to identify the language and discourse features that diagnostically predict the genre to which a text belongs (McCarthy, Myers, Briner, Graesser, & McNamara, 2009). A reader’s comprehension of a text can be facilitated by correctly identifying the textual characteristics that signal its genre (Biber, 1988). Researchers in educational psychology have shown that training struggling readers to recognize genre and global text structures helps them improve comprehension (Meyer & Wijekumar, 2007; Williams, Stafford, Lower, Hall, & Pollini, 2009). As discussed earlier, students read texts very differently if they view it as a newspaper article versus literature (Zwaan, 1994).

Identification of topic sentences in paragraphs is also an important component of rhetorical composition, at least for informational texts. These sentences are claims or main points that are elaborated by other sentences in the paragraph. They ideally occur in the *paragraph initial* position, although that ideal does not hold up in naturalistic texts (Popken, 1991). Coh-Metrix provides a number of measures of topic sentencehood that are either intrinsic characteristics of sentences or relative measures that involve comparisons between sentences in a paragraph.

Coh-Metrix has been used in dozens of projects that investigate characteristics of discourse, comprehension, memory, and learning (Crossley, Louwerse, McCarthy, & McNamara's, 2007; Graesser, Jeon, Yang, & Cai, 2007; McNamara et al., in press). These studies have validated the Coh-Metrix measures by comparing the computer output to language and discourse annotations by experts, to texts scaled on cohesion, to psychological data (e.g., ratings, reading times, memory, test performance), and to samples of texts that serve as gold standards. Coh-Metrix has uncovered differences among a wide range of discourse categories at level 5, such as differences between (a) spoken and written samples of English, (b) physics content in textbooks, texts prepared by researchers, and conversational discourse in tutorial dialogue, (c) articles written by different authors, (d) sections in typical science texts, such as *introductions*, *methods*, *results*, and *discussions*, and (e) texts that were *adopted* (or authentic) versus *adapted* (or simplified) for second language learning.

The Coh-Metrix tool should help advance theory and empirical research in discourse comprehension on a number of fronts. First, researchers can scale their texts on multiple levels of comprehension in order to perform manipulation checks, assess the impact of potential extraneous variables, and explore how the different levels of meaning are interrelated. Discourse researchers are routinely haunted that some extraneous variable may be responsible for their claims about the impact of text on psychological processes; Coh-Metrix can be used to measure and assess the potential extraneous variables. Second, researchers can test theoretical claims that texts have particular properties by collecting Coh-Metrix measures on a large corpus of naturalistic texts that are selected with rigorous scientific sampling procedures. This is a landmark advance over research 20 years ago when researchers cherry picked a handful of texts to suite their purposes. Third, researchers can discover new discourse patterns by applying data mining procedures to the hundreds of Coh-Metrix dimensions when applied to thousands of texts. We anticipate many new research breakthroughs on multilevel discourse comprehension with computation tools such as Coh-Metrix, particularly if it is used in interdisciplinary efforts with researchers in psychology, linguistics, education, language arts, communication, computer science, and many other areas of the cognitive sciences.

Acknowledgements

This research was supported by the National Science Foundation (ITR 0325428, BCS 0904909), the Institute of Education Sciences (R305G020018, R305H050169, R305B070349, R305A080589), and the US Department of Defense Counterintelligence Field Activity (H9C104-07-0014). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, and DoD. Our gratitude goes to Zhiqiang Cai for his software development of Coh-Metrix, to Phil McCarthy, Max Louwerse, and Moongee Jeon for their testing of Coh-Metrix on text samples.

References

Baker, L. (1985). Differences in standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly, 20,* 298-313.

Bard, E.G., Anderson, A.H., Chen,Y., Nicholson, H.B.M., Harvard, C., & Dalzel-Job, S.(2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language, 57,* 616-641.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, England: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Bransford, J. & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior,11*,717-796.

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,* 329-345.

Charniak, E. (2000). A maximum-entropy-inspired parser. *Proceedings of the First Conference on North American Chapter of the Association for Computational Linguistics* (pp. 132-139). San Francisco: Morgan Kaufmann Publishers.

Chi, M.T.H., de Leeuw, N., Chiu, M. & LaVancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science, 18*, 439-477.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Crossley, S.A., Louwerse, M., McCarthy, P.M., & McNamara, D.S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal, 91,* 15-30.

Daneman, M., Lennertz, T. &Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes,22*,83-105.

Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16,* 228-232.

Fellbaum, C. (Ed.). (1998). WordNet: An Electronic Lexical Database [CD-ROM]. Cambridge, MA: MIT Press.

Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Glenberg, A. M. (1997). What memory is for? *Behavior and Brain Sciences, 20,* 1-55.

Glenberg, A.M., & Kaschak, M.P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9*, 558-565.

Glenberg, A.M., & Robertson, D.A. (1999). Indexical understanding of instructions. *Discourse Processes, 28,* 1-26.

Goldman, S.R., Varma, S., & Cote, N. (1996). Extending capacity-constrained construction integration: Toward "smarter" and flexible models of text comprehension. In B.K. Britton and A.C. Graesser (Eds), *Models of understanding text* (pp. 73-114). Mahwah, NJ: Erlbaum.

Gough, P.B. (1972). One second of reading. In J.F. Kavanaugh and J.G. Mattingly Eds.), *Language by ear and by eye* (pp. 331-358). Cambridge, MA: MIT Press.

Graesser, A. C., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, *70*, 3–22.

Graesser, A. C., D'Mello, S., & Person, N. K. (2009). Metaknowledge in tutoring. In D. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education.* Mahwah, NJ: Taylor & Francis.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. (Eds.). (2003). *Handbook of discourse processes*. Mahwah, NJ: Erlbaum.

Graesser, A. C., & Jackson, G. T. (2008). Body and symbol in AutoTutor: Conversations that are responsive to the learners' cognitive and emotional states. In M. de Vega, A. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 33–56). Oxford, U.K.: Oxford University Press.

Graesser, A. C., Jeon, M., Yang, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, *15*, 199–213.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180-193.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, *33*, 1235–1247.

Graesser, A.C., & McNamara, D.S. (in press). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 193-202.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-189.

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

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, *114*, 357-374.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition in education*. Mahwah, NJ: Erlbaum/Taylor & Francis.

Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hancock, J., & Dunham.P. (2001). Language use in computer-mediated vommunication: The role of coordination devices. *Discourse Processes*, *31*, 91-110.

Hanna, J.E. & Brennan, S.E.(2007). Speakers' eye gaze disambiguates referring expressions early during face-to face conversation. *Journal of Memory and Language*,*57*,596-615.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141-158.

Holler,J.& Wilkin,K.(2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*,*24*, 267-289.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Englewood, NJ: Prentice Hall.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston: Allyn & Bacon.

Just M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kaakinen,J.K. & Hyona,J. (2007). Perspective effects in repeated reading: An eye movement study. *Memory & Cognition, 35*, 1323-1336.

Kendeou, P. & van den Broek, P. (2007). The effects of prior knowledge in text structure on comprehension processes during reading scientific texts. *Memory & Cognition, 35,* 1567-1577.

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

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

Klare, G. R. (1974–1975). Assessing readability. *Reading Research Quarterly*, *10*, 62-102.

Koslin, B.I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading*. New York: College Entrance Examination Board.

Kosslyn, S.M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007)(Eds.). *Handbook of Latent Semantic Analysis.* Mahwah, NJ: Erlbaum.

Lappin, S., & Leass, H. J. (1994). An algorithm for pronominal coreference resolution. *Computational Linguistics*, *20,* 535-561.

Louwerse, M.M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*, 291-315.

Magliano, J.P., & Radvansky, G.A. (2001). Goal coordination in narrative comprehension. *Psychonomic Bulletin and Review, 8*, 372-376.

Magliano, J., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.

Mandler, J.M. (1988). How to build a baby: On the development of an accessible representational system. *Cognitive Development*, 3, 113-136.

Masson, M.E.J., Bub, D.N., & Warren, C.M.(2008).Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language, 59,* 256-265.

McCarthy, P. M., Myers, J. C., Briner, S. W., Graesser, A. C., & McNamara, D. S. (2009). Are three words all we need? A psychological and computational study of genre recognition. *Journal for Language Technology and Computational Linguistics, 1,* 23-57.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38,* 1-30.

McNamara, D.S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-287.

McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (in press). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*.

McNamara, D.S., & Magliano, J. (2008). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The Psychology of Learning and Motivation,* Vol. 51.

Meyer, B. J. F., & Wijekumar, K. (2007). Web-based tutoring of the structure strategy: Theoretical background, design, and findings. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (347-375). Mahwah, NJ: Lawrence Erlbaum Associates.

Millis, K.K., Kim, H.J., Todaro, S. Magliano, J., Wiemer-Hastings, K., & McNamara, D.S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers, 36*, 213-221.

Millis, K.K., King, A., & Kim, J. (2001). Updating situation models from descriptive texts: A test of the situational operator model. *Discourse Processes, 30*, 201-236.

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.

Nygaard, L.C.,Herold, D.S,. & Namy,L.L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*,*33*,127-146.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1200-1210.

O'Reilly,T. & McNamara, D.S.(2007a). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement.*American Educational Research Journal, 44,*161-197.

O'Reilly, T. & McNamara, D.S. (2007b). Reversing the reverse cohesion effect: good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43,* 121-152.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what the read. *Psychological Science, 3,* 229-235.

Padó,U., Crocker, M.W., & Keller,F.(2009). A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science*,*33*,794-838.

Pecher,D. & Zwaan, R.A.(2005). *Grounding cognition: The role of perception and action in memory,language, and thinking.* Cambridge, England: Cambridge University Press.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count*. Austin, TX: LIWC.net (www.liwc.net).

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*,*11*, 357-383.

Pickering, M., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.

Pollatsek,A., Slattery,T.& Juhasz,B.(2008). The processing of novel and lexicalised prefixed words in reading. *Language and Cognitive Processes*,*23*,1133-1158.

Popken, R. (1991). A study of topic sentence use in technical writing. *The Technical Writing Teacher, 18*, 49-58.

Rapp, D.N. (2008). How do readers handle incorrect information during reading?*Memory & Cognition, 36,* 688-701.

Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rayner, K., & Pollatsek, A. (1994). *The psychology of reading*. Mahwah, NJ: Erlbaum.

Riesco-Bernier,S.&Romero-Trillo, J.(2008). The acoustics of "newness" and its pragmatic implications in classroom discourse. *Journal of Pragmatics*,*40*,1103-1116.

Rouet, J. (2006). *The skills of document use: From text comprehension to web-based learning.* Mahwah, NJ: Erlbaum.
Rumelhart, D.E. (1977). Toward an interactive model of reading In S. Dornie (Ed.), *Attention and performance* (pp. 573-603). Hillsdale, N. J.: Erlbaum
Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*, 37–60.
Sanford, A.J., & Graesser, A.C. (2006). Introduction: Shallow processing and underspecification. *Discourse Processes, 42,* 99-108.
Schober, M.F. & Brennan, S.E. (2003). Processes of interactive spoken discourse. In A.C. Graesser, M.A. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp 123-164). Mahwah, New Jersey: Erlbaum.
Sinclair, J.M., & Coulthard, R.M. (1975). *Towards an analysis of discourse: The English used by teachers and their pupils.* London: Oxford University Press.
van der Sluis,L.& Krahmer,E. (2007). Generating multimodal references. *Discourse Processes, 44*, 145-174.
Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.
Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407.
Stenner, A.J.(2006). *Measuring reading comprehension with the Lexile framework*. Durham, NC: Metametrics, Inc. presented at the California Comparability Symposium, October 1996. Retrieved January 30, 2006 from http://www.lexile.com/DesktopDefault.aspx?view=re.
Sturt, P., Sanford, A.J., Stewart, A.J., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change detection paradigm. *Psychonomic Bulletin and Review, 11*, 882-888.
Taraban, R. & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory & Language, 27*, 597-632.
Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain: Long-term memory encoding. *Psychonomic Bulletin & Review, 2,* 387-390.
Van den Broek, P., Virtue, S., Everson, M.G., & Tzeng, Y., & Sung, Y. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 131-154). Mahwah, NJ: Erlbaum.
Van den Broek, P., Rapp, D.N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. Discourse Processes, 39, 299-316.
VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., & Rose, C.P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31,* 3-62.
van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic
De Vega, M., Glenberg, A. M., & Graesser, A. C. (Eds.). (2008). *Symbols and embodiment: Debates on meaning and cognition*. Oxford, U.K.: Oxford University Press.
Ward, P. & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition, 35,* 73-86.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (in press). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*.

Williams, J.P., Stafford, K.B., Lauer, K.D., Hall, K.M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*, 1-20.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *20*, 920-933.

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*, 162-185.

Table 1: *Levels of Language and Discourse*

LEVEL	EXAMPLE COMPONENTS OF LEVEL
Words	Lexical meaning representation Word composition (graphemes, phonemes, syllables, morphemes, lemmas) Parts of speech (noun, verb, adjective, adverb, determiner, connective)
Syntax	Syntax (noun-phrase, verb-phrase, prepositional phrase, clause) Linguistic style
Textbase	Semantic meaning Explicit propositions or clauses Referents linked to referring expressions Bridging inferences that connect propositions, clauses, or words Connectives that explicitly link clauses
Situation model	Situation conveyed in the text Agents, objects, and abstract entities Dimensions of temporality, spatiality, causality, intentionality Inferences that elaborate text and link to the reader's experiential knowledge Given versus new information Images and mental simulations of events
Genre & rhetorical structure	Discourse category (narrative, persuasive, expository, descriptive) Rhetorical composition (cause+effect, claim+evidence, problem+solution) Epistemological status of propositions and clauses (claim, evidence, warrant) Speech act categories (assertion, question, command, request, greeting, etc.) Theme, moral, or point of discourse
Pragmatic communication	Goals of author Attitudes and beliefs (humor, sarcasm, eulogy, deprecation)

Table 2: *Example Coh-Metrix Measures and Indices (over 700 available)*

LEVEL OR CLASS	MEASURE (INDEX)
Words	Frequency, concreteness, imagery, age of acquisition, part of speech, content words, pronouns, negations, connectives (different categories), logical operators, polysemy, hypernym/hyponym (reflects abstractness); these counts per 1000 words.
Syntax	Syntactic complexity (words per noun-phrase, words before main verb of main clause)
Textbase cohesion	Cohesion of adjacent sentences as measured by overlapping nouns, pronouns, meaning stems (lemma, morpheme). Proportion of content words that overlap. Cohesion of all pairs of sentences in a paragraph.
Situation model cohesion	Cohesion of adjacent sentences with respect to causality, intentionality, temporality, spatiality, and latent semantic analysis (LSA). Cohesion among all sentences in paragraph and between paragraphs via LSA. Given versus new content.
Genre and rhetoric	Type of genre (narrative, science, other). Topic sentencehood
Other	Flesch-Kincaid grade level, type token ration, syllables per word, words per sentence, sentences and paragraphs per 1000 words.