Inference Generation and Cohesion in the Construction of Situation Models:
Some Connections with Computational Linguistics

Arthur C. Graesser, Max M. Louwerse, Danielle S. McNamara, Andrew Olney,
Zhiqiang Cai, and Heather H. Mitchell

The University of Memphis


Send correspondence to:

Art Graesser
Department of Psychology
202 Psychology Building
University of Memphis
Memphis, TN 38152-3230
(901) 678-2742
(901) 678-2579 (fax)
a-graesser@memphis.edu
Inference Generation and Cohesion in the Construction of Situation Models:

Some Connections with Computational Linguistics

It is widely accepted in discourse psychology that readers construct situation models (mental models) as they attempt to comprehend text (Graesser, Millis, & Zwaan, 1997; Kintsch, 1998; Zwaan & Radvansky, 1998). When the text is in the narrative genre, such as a simple story, the situation model is a mental microworld of events and actions, with characters performing actions in the pursuit of goals, events that present obstacles to the goals, conflicts between characters, methods of resolving conflicts, and emotional reactions to the events and conflicts. If the reader has sufficient time and motivation, the mental microworld can have quite vivid elaborations that flesh out details of the spatial setting, the style and procedure of actions, and the properties of objects that characters use. There may also be content that refers to mental states of characters (what they believe, know, perceive, and want). Mental models are created from information in the text as well as knowledge of the reader.

It is comparatively easy for adults to construct situation models for narrative texts because the microworlds have a close correspondence to everyday experiences. In contrast, it is more difficult to construct a situation model for informational texts, such as an expository text on a scientific mechanism. Ideally, the reader would be able to construct a mental model of the components, causal chains, and processes that capture a scientific mechanism. However, this is very difficult or impossible when the reader has little world knowledge to furnish the content needed to construct the mental model.

Discourse psychologists also widely acknowledge that the explicit text does not sufficiently constrain or determine what information is constructed in the situation model. Instead, the content in the situation model is the result of complex interactions among (a) explicit
features in the text, (b) capacities of the reader (such as background world knowledge, generic reading skills, cognitive limitations), and (c) the task or goals the reader is attempting to achieve while reading the text (Kintsch, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996; Snow, 2002). Thus, it is appropriate to view comprehension as a complex function or mechanism that considers Representation x Text x Reader x Goal interactions. An explanatory account of situation model construction has required cognitive models that are far more sophisticated and detailed than a handful of general laws or principles. Prominent examples of these cognitive models in discourse psychology include the construction integration model (Kintsch, 1988, 1998; Schmalhofer, McDaniel, Keefe, 2002; Singer & Kintsch, 2001), the constructionist theory (Graesser, Singer, & Trabasso, 1994; Singer, Graesser, & Trabasso, 1994), the structure building framework (Gernsbacher, 1997), the event indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998), memory-based resonance models (Lorch, 1998; O’Brien, Rizzella, Albrecht, & Halleran, 1998), and the landscape model (van den Broek, Everson, Virtue, Sung, & Tzeng, 2002).

Research on discourse comprehension will presumably grow in sophistication (and hopefully not confusion or obfuscation) to the extent that there is more analytical detail and a more multidisciplinary perspective in specifying the representations and processes. A scientific understanding of deeper levels of comprehension has benefited from several fields that offer enriched analytical detail, as history has proven (Graesser, Gernsbacher, & Goldman, 2003). In the 1970’s, the fields of linguistics and text linguistics identified the constituents of explicit sentences and text (van Dijk, 1972); the field of artificial intelligence identified the representational formats and structures of different types of world knowledge (Schank & Abelson, 1977; Winograd, 1972); cognitive psychology identified the processes that access, use,
and manipulate the text representations (Bower, Black, & Turner, 1979; Kintsch, 1974). In the 1980’s and 1990’s, there was a shift to more complex and dynamic processing architectures that dominated the fields of cognitive science, such as production systems (Just & Carpenter, 1992) and neural networks (Kintsch, 1988; St. John, 1992). Graesser et al. (2003) have forecasted that the next generation of discourse psychologists will substantially benefit from contributions in corpus linguistics, computational linguistics, and neuroscience.

In this chapter, we hope to achieve two objectives. The first objective will be to briefly identify a landscape of inferences and cohesion relations. Researchers in discourse psychology are encouraged to consider these if they want a comprehensive vision of comprehension. It will prevent researchers from overly concentrating on a small set of inferences and relations, at the expense of ignoring a large chunk of the terrain. A portion of this landscape of inferences and cohesion relations has already been investigated by discourse psychologists who collect behavioral data, but there still are regions that have been neglected.

A second objective of this chapter is to briefly identify some advances in computational linguistics that offer promising ideas and tools for discourse psychology and cognitive neuroscience. In these advances, researchers have managed to automate aspects of language and comprehension modules, which is no small engineering feat. For example, as we will discuss in this chapter, we have made some headway in our development of Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) Coh-Metrix analyzes texts on several dozens of measures of cohesion and language. One version of Coh-Metrix under development takes into consideration the reader’s reading skills and prior knowledge. The contributions of computational linguistics, computational discourse, and Coh-Metrix extend beyond engineering and into science to the extent that the representations and processing modules have
correspondences to the components in our theories of cognition and neuroscience. Whether these correspondences succeed or fail very much remains an open question.

Inferences

Table 1 lists and describes 13 classes of inferences that are potentially constructed during text comprehension. This classification was adopted from Graesser, Singer, and Trabasso (1994). The classification scheme covers a diverse set of inferences, although one might question whether this list is exhaustive and whether the grain size and categorical distinctions are optimal for conducting a program of research. All of these classes of inferences require world knowledge to construct. These knowledge-based inferences are constructed by activating and recruiting world knowledge in an effort to make sense out of the text. They are not constructed by logical truth tables, predicate calculus, propositional calculus, Bayes theorem, and statistical algorithms that many researchers and scholars have traditionally associated with inferences (Kahneman, Slovic, & Tversky, 1982; Rips, 1983; Russell & Whitehead, 1925).

INSERT TABLE 1 ABOUT HERE

It is beyond the scope of this chapter to define these inference classes precisely and to review the empirical research that assesses the extent to which they are generated during comprehension. However, we can offer a few observations. The classes of inferences differ according to the span of text that is needed to generate the inferences. Classes 1-3 are normally triggered by single words, classes 4-9 by clauses or sentences, and classes 12-13 by lengthy stretches of text, if not the entire composition. It is easiest, methodologically, to investigate inferences that are triggered by shorter spans of text because the researcher can readily designate the points in the text when the inferences are supposed to be made. Most of the available research on on-line comprehension has been conducted on inference classes 1, 4, 5, 6, 7, 8, and
11, whereas there has been a modest amount of research on classes 2, 3, 9, and 10, and very little research on classes 12 and 13. When evaluating the status of an inference with respect to online processing, researchers have used a variety of behavioral measures and methods: self-paced reading times for words or sentences, gaze durations in eye tracking studies, latencies of word naming and lexical decision judgments for test words that periodically occur while reading text, speeded recognition judgments for test words or sentences, and so on (Haberlandt, 1994). There recently has been a number of studies that have used methods in neuroscience, such as evoked potentials and fMRI (Beeman, Bowden, & Gernsbacher, 2000; Griesel, Friese, & Schmalhofer, 2003; Reichle, Carpenter & Just, 2000; Wharton et al., 2000).

The classes of inferences presumably differ in how readily they are generated on line. For the sake of the present chapter, the processing status of an inference may be viewed as being in one of a handful of categories. Inferences are automatic if they are generated reliably and very quickly (i.e., within a half a second, approximately, of the onset of an inference triggering word), with very little cognitive effort and processing resources. Inferences are routine if they require more cognitive effort, but are made reliably and moderately quickly (within a second or less). Inferences are strategic if their construction is sensitive to the reader’s goals and comprehension strategies, but otherwise are generated moderately quickly. Inferences are off-line if they can only be made after comprehending the text and ordinarily with considerable time, effort, and perhaps extended study and reflection. If we had to place our bets, examples of automatic, routine, strategic, and off-line inferences would be classes 2, 4, 11, and 13, respectively. That is, an example of an automatic inference is the case structure role assignment, where the reader automatically infers the functional role of each explicit noun-phrase or prepositional phrase (e.g., agent, recipient, object, time, location). Routine inferences are less automatic, as in the case
superordinate goal inferences. The reader needs some additional time to access knowledge structures from long-term memory that help reconstruct the inferred goals of explicit intentional actions. An example of a strategic inference is a state that specifies the location of an agent or object in a spatial layout. Details about spatial location of entities are not normally constructed on-line unless there is an explicit goal to monitor spatiality and there is enough time to construct such detail. The author’s intention is an example of an inference that ordinarily is constructed off-line, after some reflection.

Although this classification of inferences provides a reasonably accurate overview of inferences during comprehension, there are some important caveats. In any test of this classification, it is essential that the sample of readers have the prerequisite amount of world knowledge and reading skill when offering a fair test of the processing status of an inference. All bets are off if the readers lack critical world knowledge. As a second caveat, it is important that the composition of the text is sufficiently well-formed when offering a fair test of the processing status of an inference. All bets are off if the text is pointless, incoherent, haphazardly assembled, or otherwise “inconsiderate.” For example, it would be inappropriate to conclude that theme inferences are not made if the text in a study had a scrambled sequence of experimenter-generated sentences. These two caveats provide important lessons for discourse psychology, computational linguistics, and cognitive neuroscience. Researchers must consider the background knowledge and reading skills of the reader before firm conclusions can be made about the processing status of inferences. Although experimenter-manipulated texts allow for control over potential extraneous variables, they introduce a set of confounding variables that accrue from unnaturalness. Without some assessment of a phenomenon on a natural text, the validity of the phenomenon is fundamentally threatened.
The classification scheme in Table 1 was proposed when the constructionist theory was introduced by Graesser et al. (1994). Among other things, the constructionist theory attempted to predict and explain what classes of inferences are routinely made during comprehension and what classes are not. There were two distinctive assumptions of the constructionist theory: the coherence assumption and explanation assumption. The coherence assumption states that readers attempt to construct a meaning representation that is coherent at both local and global levels. Local coherence refers to the content, structures, and processes that organize elements, referents, and constituents of adjacent clauses or short sequences of clauses. Global coherence is established when these local chunks of information are organized and interrelated into higher order chunks. The explanation assumption states that readers attempt to explain why actions, events, and states are mentioned in a text. The explanations normally include naïve theories of physical and psychological causality, but the content would have scientific integrity to the extent that the reader has more background knowledge in science.

The other assumptions of the constructionist theory are less distinctive in the sense that they are frequently adopted by other theories of comprehension. For example, the reader goal assumption states that readers construct inferences that address their comprehension goals and the task that they are attempting to achieve. Other uncontroversial components and assumptions stipulated that there are multiple information sources, different levels of representation (surface code, textbase, situation model), multiple memory stores (short-term, long-term, and working memory), a discourse focus, the notion that the strength of encoding an idea increases by activation from multiple sources of information (called convergence) and constraint satisfaction, and the notion that elements in a knowledge structure are more accessible to the extent that there is repetition and automaticity from past experiences.
It is the coherence assumption and the explanation assumptions that made the most discriminating predictions about the classes of inferences that are routinely made during comprehension. According to the coherence assumption, the classes that would have the processing status of being routine or automatic would be referential (1), case structure role assignment (2), causal antecedent (7), and theme (12). These inferences are needed to link together constituents of the explicit text. According to the explanation assumption, the routine or automatic inferences would be superordinate goals (4) and causal antecedents (7). These inferences are generated as answers to why-questions. The remaining categories of inferences would be strategic, off-line, or contingent on the outcome of the uncontroversial components and assumptions. Graesser et al. (1994) reviewed the empirical evidence for the constructionist theory and identified the predictions of alternative theoretical positions. It is a matter of debate how well the constructionist theory has survived after a decade of empirical research and a number of new theoretical models that have entered the arena, such as memory-based models (Lorch, 1998; O’Brien et al., 1998), the landscape model (van den Broek et al., 2002), and embodied theories of cognition (Glenberg, 1997; Kaup, Zwaan, & Lüdtke, this volume, 2005; Zwaan, Stanfield, & Yaxley, 2002).

It is informative to illustrate how the constructionist theory would generate predictions different from alternative theoretical predictions. Consider the following excerpt from the beginning of Ian Rankin’s 1992 novel A Good Hanging.

It was the perfect Murder.

Perfect, that is so far as the Lothian and Borders Police were concerned. The murderer had telephoned in to confess, had then panicked and attempted to flee, only to be caught leaving the scene of the crime. End of story. (p. 1)
The following four inferences would potentially be made at particular points in the text.

(A) **Superordinate goal**: “The murderer wanted to turn himself into the police” is an inference when reading the clause *The murderer had telephoned in to confess.*

(B) **Subordinate goal/action**: “The murderer dialed a telephone number” is an inference when reading the clause *The murderer had telephoned in.*

(C) **Causal antecedent**: “The murderer changed his mind” is an inference when reading the clause *he then panicked.*

(D) **Causal consequence**: “The murderer escaped” is an inference when reading the clause *he then panicked.*

According to the constructionist theory, adult readers would generate A and C as routine or automatic inferences, but not B and D. Inferences A and C are answers to why-questions, the signature question for explanation-based reasoning. So when asked “Why did the murderer telephone in?”, a plausible answer would be “in order to turn himself in to the police.” When asked “Why did the murderer panic?”, a plausible answer would “because he changed his mind and did not want to turn himself in to the police.” Inference B is an answer to a how-question (“How did the murderer telephone in?”), not a why-question, so it would be mere ornamentation and strategically generated rather than being a routine inference. Inference D is an answer to a what-happened-next-question (“What happened after the murderer panicked?”), not a why-question.

The predictions are quite different for other theoretical frameworks. The minimalist hypothesis of McKoon and Ratcliff (1992) would predict that none of the inferences would be generated, whereas the script theory of Schank and Abelson (1977) would predict that all four of the inferences would be generated. Some theoretical models would offer predictions for some of
these inference classes but not other classes. One can track each inference class, and derive predictions for each theoretical position. For example, causal consequences (D) are predicted to be routine inferences by a prediction-substantiation model (DeJong, 1979) and script theory, but not routinely by Kintsch’s construction integration model, the minimalist hypothesis and the constructionist theory. Subordinate goals/actions (B) are predicted to be routine inferences by script theory and an embodied theory of cognition (Glenberg, 1997; Zwaan et al., 2002), but not by the minimalist hypothesis and the constructionist theory. It is important to acknowledge that we have specified all-or-none predictions for these positions for the sake of simplicity; clearly, these inferences vary in strength of activation rather than being generated all-or-none. For example, the constructionist theory does specify special conditions when causal consequence and subordinate goal/action inferences are generated (Graesser et al., 1994), whereas Schmalhofer, McDaniels, and Keefe (2002) have augmented Kintsch’s CI model to accommodate the possibility of the prediction-substantiation mechanism. Nevertheless, the predictions that we have outlined offer testable claims about the relative strength and reliability of generating such inferences.

Predictions of the constructionist theory were tested in several empirical studies published in the 1990’s. These studies collected word naming latencies and lexical decision latencies for test words that were presented quickly after sentences were read in the text (Long, Golding, & Graesser, 1992; Magliano, Baggett, Johnson, & Graesser, 1993; Millis & Graesser, 1994), reading times for sentences in text (Graesser & Bertus, 1998), and think aloud protocols (Magliano, Trabasso & Graesser, 1999). However, empirical tests in the arena of neuroscience are sparse, so this is one direction for future research. The classes of inferences presumably differ with respect to the areas of the brain that are activated and with respect to patterns of brain
activity. For example, Griesel, Friese, and Schmalhofer (2003) reported that bridging inferences (i.e., categories 1 and 7 in Table 1) are accessed more rapidly in the left hemisphere whereas predictive inferences (categories 4 and 8) are accessed more rapidly in the right hemisphere. There is some evidence that the predictive inferences start out being generated in the right hemisphere before being represented in the left hemisphere later on (Beeman et al., 2000). Brain imaging technologies (such as fMRI and ERP’s) can be used to better inform us on the processing status of the different classes of inferences.

Cohesion Signals and Relations

A fundamental distinction can be made between coherence and cohesion. Coherence is a characteristic of the representation in the mind of the reader, whereas cohesion is a characteristic of the text. A reader perceives a text to be coherent to the extent that the ideas conveyed in a text hang together in a meaningful and organized manner. Coherence relations are constructed in the mind of the reader if the reader has adequate world knowledge about the subject matter and if there are adequate linguistic and discourse markers. Thus, coherence is an achievement that is a product of psychological representations and processes. In contrast, cohesion is an objective property of the explicit language and discourse. There are explicit features, words, phrases, or sentences that guide the reader in interpreting the substantive ideas in the text, in connecting ideas with other ideas, and in connecting ideas to higher level global units (e.g., topics, themes). Simply put, coherence is a psychological construct whereas cohesion is a textual construct (Louwerse, 2002; Louwerse & Graesser, in press).

Graesser, McNamara, and Louwerse (2003) identified several categories of cohesion signals and relations that exist in narrative and informational text. These cohesion signals and relations are summarized in Table 2. As in the case of inferences, this list of categories is not
intended to be exhaustive or to be the theoretically most natural classification. Instead, it provides a representative landscape of what researchers need to worry about. As one moves from class 1 to 10, there is a general shift in the span of the text units (and the associated situation model) from local to global. That is, the span grows from a single word or phrase (classes 1 and 2), to a single clause or sentence (classes 3 and 4), to pairs or small sets of clauses or sentences (classes 5, 6, and 7), to paragraphs (class 8), and to sections of text (classes 9 and 10).

One of the interesting challenges for discourse researchers is to understand the coordination of text cohesion with a coherent situation model. There are some advantages to having a tight coupling between (a) cohesion signals and relations and (b) the ideal coherent situation model that the author intends to convey. For example, if the ideal situation model is a causal chain of four events (A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D), then a considerate text would present the events in the chronological order, with the same verb tense, and with causal conjunctive relations that connect some of the adjacent event pairs (such as those event pairs that some readers might not believe are causally related in the absence of a causal connective). Indeed, a tight coupling between cohesion and the situation model is desirable for low knowledge readers because they rely on cohesion markers to direct their deeper comprehension (Britton & Gulgoz, 1991; McNamara, 2001; McNamara et al., 1996).

On the other hand, it is not practical and sometimes not even desirable to have a tight coupling between the cohesion markers and the situation model. Consider the following five events that describe part of the mechanism of a cylinder lock, as described in Macaulay’s *The Way Things Work* (1988).
Text 0:

(A) The key turns.

(B) The cam rotates.

(C) The lip of the cam pulls back the rod.

(D) The bolt that is connected to the rod moves into the door.

(E) The door opens.

In the original text, the five events in text 0 were accompanied by a picture and some surrounding text that made the comprehension experience flow more naturally; these five events are extracted here for the sake of making some points. One point is that it would be cumbersome if there were a relational, causal verb (such as cause, enable, allow) that connected each pair of events, as illustrated below in text 1.

Text 1:

The key’s turning causes the cam to rotate. The rotation of the cam causes the lip of cam to pull back the rod. The pulling back of the rod by the lip of the cam causes the bolt that is connected to the rod to move into the door. The bolt’s moving into the door allows the door to open.

Although text 1 has a distinct advantage of coreferential cohesion, the text is undesirable in several respects. The sentences are longer, sometimes to the point of potentially overloading working memory. The syntactic composition of some sentences is awkward, as in the case of the third sentence, which results in potential comprehension difficulties. There is a need to create gerund expressions (the key’s turning) and nominalizations (the rotation of the cam) in order to accommodate the causal verb and also to specify the events. Gerunds and nominalizations are syntactically dense constructions that might challenge readers with low reading ability. There is
added redundancy because events B, C, and D need to be mentioned twice. This has the virtue of repetition, which is presumably helpful to learning, but has the liability of wordiness. Perhaps the repetition can be avoided by connecting only some of the adjacent pairs with causal verbs. However, that approach would require a decision on which adjacent pairs to connect and would also have a looser coupling between cohesion and the situation model.

Text 2 articulates the same five events with conjunctive relations that specify time or causality.

Text 2:

(2) When the key turns, the cam rotates. As a result, the lip of the cam pulls back the rod. Consequently, the bolt that is connected to the rod moves into the door, so the door can open.

Despite the sacrifice of some referential cues, text 2 is clearly a much smoother and concise articulation than text 1. However, there are a few difficulties with the use of conjunctive relations as a general solution to the specification of causal relations between events. One problem is that most of the conjunctions frequently used in English have undesirable features. Some of the conjunctions are not necessarily causal, such as when, and then and so. When and and then may merely convey temporality, whereas so may suggest a logical derivation rather than causality (Louwerse, 2002). Such polysemous conjunctions fail to discriminatively pinpoint the causal relation. Consequently is patently causal, although it is debatable what sort of causal relation is implied (e.g., direct or indirect cause). Because is the prototypical causal conjunction, but has the undesirable feature of reversing temporal order of events in articulation frames X because Y. For example, it is inappropriate to say the key turns because the cam rotates (A because B), even though the order of mentioning the events follows the chronological causal
order in the situation model. Instead, it is appropriate to say *the cam rotates because the key turns* (B because A), where the order of mentioning events is opposite to the chronological causal order.

Writers often remove information in order to cope with the complexity of mapping cohesion markers onto the intended situation model. This is accomplished by deleting some of the causal relations and by using noun-phrases that elliptically delete fragments of the event being referred to. This is illustrated in Text 3.

**Text 3:**

(3) When the key turns, the cam rotates. The rotation causes the lip of cam to pull back the rod. Then the bolt moves into the door so the door can open.

*The rotation* in the second sentence elliptically deletes mentioning that it is the cam that rotates. There is no causal link between the pulling back of the rod and the bolt moving into the door, and no mention of the fact that the bolt is connected to the rod. Unfortunately, the removal of this content will likely create problems for low knowledge readers because they will be unable to fill in the missing information and construct a coherent causal thread. The sentences are indeed shorter, which presumably alleviates working memory load and creates simpler syntactic constructions; both of these are allegedly bonuses for readers with low reading skill. However, the liability is that the integrity of the situation model suffers.

McNamara has documented an intriguing interaction between text cohesion and the world knowledge of the readers when they attempt to comprehend science texts (McNamara, 2001; McNamara et al., 1996). In a series of studies, McNamara manipulated text cohesion (high versus low), measured the prior knowledge that readers had about the topics (high versus low knowledge about the science topic), and administered tests that assessed different levels of
meaning representation (recall of the explicit propositions versus answering questions about the deeper situation model). High-cohesion texts were best for low-knowledge readers, no matter what type of test was administered. This unsurprising result is compatible with most theories of comprehension and the folklore of teachers and text designers. Recall for high-cohesion texts was also a bit higher than, or equivalent to, low cohesion texts for readers with high world knowledge. Again, this outcome is not particularly surprising. However, in tests of the deeper situation model, high-knowledge readers frequently benefited from text with low cohesion. The low-cohesion texts encouraged the knowledgeable readers to work harder and build more elaborated situation models. This result strongly suggests that it is important to tailor the cohesion of the text to the world knowledge of the reader. High knowledge readers may benefit from more challenges and difficult texts in order to prevent them being lulled into complacency from a well-crafted text that engenders an illusion of comprehension. Low knowledge readers benefit from a high density of coherence relations.

It would be intriguing to test whether a Knowledge x Cohesion interaction occurs in evoked potentials, fMRI’s, or other paradigms in cognitive neuroscience. The experimental procedure would be quite straightforward in the case of fMRI. Participants would be administered tests of individual differences on world knowledge, reading fluency, language decoding, working memory, and other measures of verbal ability (Perfetti, 1985, 1994). As they read texts that vary in cohesion, researchers would collect fMRI data and observe the extent to which areas of the brain are active for theoretically expected dimensions of the situation model. For low-knowledge readers, brain activities in these areas should be more pronounced for high-cohesion than low-cohesion texts; the opposite trend should occur for high knowledge readers. Of course, a fair test of the impact of cohesion should rule out auxiliary dimensions of language
and the situation model. The computer tool described in the next section scales the texts on metrics that correspond to these potentially correlated dimensions.

Cohesion markers need to be systematically coordinated with the situation model, so it is extremely important to have a sufficiently detailed theoretical specification of the situation model. The content, format, and structure of the situation model are matters of substantial theoretical debate, but there is widespread agreement that the situation model is a multithreaded construct rather than a monolithic construct. This notion is emphasized in Zwaan’s event indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). According to this model, the reader monitors five conceptual dimensions during reading: The protagonist (agency), temporality, spatiality, causality, and intentionality (i.e., character goals). A break in continuity may occur on any one of these dimensions while reading an event $E_N$ in incoming sentence and relating it to the event $E_{N-1}$ in the previous sentence. **Protagonist** discontinuity occurs when event $E_N$ has a character that is different from the characters in the previous event $E_{N-1}$. **Temporal** discontinuity occurs when event $E_N$ occurs much later in time than $E_{N-1}$, or there is a flashback. **Spatial** discontinuity occurs when the spatial setting of $E_N$ is different from that of $E_{N-1}$. **Causal** discontinuity occurs when event $E_N$ is not causally related to $E_{N-1}$. **Intentional** discontinuity occurs when the event $E_N$ is part of a protagonist’s plan that is different from the plan in the local discourse context. An incoming event in a text may have discontinuities on more than one of these five dimensions. Zwaan, Magliano, and Graesser (1995) reported that reading time for an explicit event in a literary story increased as a function of the number of dimensions with discontinuities and that each dimension had its unique impact on reading time. Discontinuities on these dimensions also
predicted the extent to which pairs of events were associatively related, with weaker associations for pairs with more discontinuities (Zwaan, Langston, & Graesser, 1995). The neuroscience literature should shed some light on whether discontinuities and the resulting inferences on the multithreaded situation model will systematically predict brain activities. Researchers in cognitive neuroscience have examined differences in the roles of the right and left hemispheres during comprehension at different levels of processing (Ferstl, Guthke & von Cramon, 2002). Long and her colleagues (Long & Baynes, 2002; Long, Bayes, & Prat, 2005) reported that the propositional representation and the establishment of local coherence resides in the left hemisphere, whereas many aspects of the situation model and global discourse coherence reside in both hemispheres. Beeman et al. (2000) reported differences in the types of inferences processed in the two hemispheres. Participants listened to stories that stimulated inferences and then named inference-related words. The test words were presented to the right visual field (i.e., processed by the left hemisphere) or the left visual field (processed by the right hemisphere) and word priming effects were measured. The inferences needed for establishing cohesion showed priming when processed in the left hemisphere, whereas the predictive inferences had more involvement of the right hemisphere. Aside from the hemispheric differences, neuroimaging studies have suggested that the frontal cortex has a prominent function in creating coherence. Wharton et al (2000) used fMRI to investigate causal inferences that are needed to fill causal discontinuities; the bilateral dorsomedial frontal cortex was primarily involved in the construction of these causal inferences.

Individual differences will no doubt present a somewhat more complex picture of the brain activities associated with inferences, cohesion, and coherence. St. George, Mannes, and Hoffman (1997) used an electroencephalogram (EEG) to measure working memory capacity in
readers. The inferences generated during reading were affected by working memory capacity. Readers with high working memory span were able to make both bridging inferences (needed to establish text coherence) and elaborative inferences (not needed for text coherence). Readers with low working memory span were able to make only the bridging inferences.

Many of the cognitive neuroscience studies have unfortunately been imprecise in specifying the classes of inferences and cohesion relations that were used in the text materials. One direction for future neuroscience research is to be somewhat more specific about text and content features, perhaps to the level of grain size shown in Tables 1 and 2.

**CohMetrix: A Software Tool that Assesses Texts on Cohesion and Language**

Using Advances in Computation Linguistics

Recent advances in the field of computational linguistics have made it more feasible to automate the processing of language and text comprehension (Allen, 1995; Jurafsky & Martin, 2000). As a consequence of these advances, we can go some distance in accounting for discourse cohesion by identifying language and discourse patterns that are sufficiently well specified that they can be extracted by computers. We have recently developed a web-based software tool, called Coh-Metrix, that analyzes texts on hundreds of measures of cohesion, language, and readability (Graesser, McNamara et al., 2004). The ultimate goal is to have a tool that replaces standard readability formulas by being sensitive to a range of cohesion relations, classes of inferences, and reader abilities (i.e., world knowledge). This section will briefly describe Coh-Metrix and will give examples of its metrics for some example texts.

We view Coh-Metrix as an important integration of research in discourse psychology and research in computational linguistics. In the past, researchers in computational linguistics were only minimally aware of contributions in discourse psychology. Computational linguists were
primarily concerned with building automated natural language technologies that accurately process text and that incorporate linguistic theories and or statistical algorithms; there was no direct concern with cognitive representations, psychological mechanisms, and neuroscience – the concerns of the discourse psychologist. At this point, the integration of these fields is so much at its infancy that we can only raise the open question of which modules in Coh-Metrix are psychologically plausible and which are merely engineering feats.

As we have discussed in this chapter, inference generation, cohesion, and coherence are fundamental to text comprehension. However, these components are not captured in common text readability formulas that focus almost exclusively on word length, word frequency, and sentence length. Coh-Metrix measures texts on a broad profile of language and cohesion characteristics. Consequently, we anticipate that Coh-Metrix will eventually replace the current readability formulas. Another salient feature of Coh-Metrix, particularly the future versions that we develop, is that it does not predict readability of the text on the basis of textual characteristics alone. Instead, it predicts how readable a text will be on the basis of its cohesion in combination with reader characteristics (such as world knowledge and general reading ability).

Coh-Metrix (http://csep.psyc.memphis.edu/cohmetrix) analyzes texts on over 50 types of cohesion relations and over 200 measures of language and discourse. It does so through modules that use lexicons, classifiers, syntactic parsers, shallow semantic interpreters, conceptual templates, latent semantic analysis, and other components that are widely used in computational linguistics (Allen, 1995; Jurafsky & Martin, 2000; Landauer, Foltz, & Laham, 1998). For example, Coh-Matrix taps into a large number of lexicons that are available for free, including WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), MRC Psycholinguistic Database (Coltheart, 1981), and the word frequency statistics collected by Francis and Kucera (1982).
These lexicons collectively provide information about a word’s syntactic class(es), semantic composition, alternative senses, concreteness, imagability, frequency of usage of words in the English language, and dozens of other characteristics. One module uses syntactic parsers (Abney, 1997; Sekine & Grishman, 1995) and a part-of-speech “tagger” (e.g., classifier) developed by Brill (1995). Many words can be assigned to more than one part of speech (e.g., bank can be an adjective, noun, or verb) so there needs to be some way of determining what syntactic class is relevant in the sentence context; these natural language technologies automatically assign the appropriate syntactic class. The syntactic parser also provides the foundation for scaling sentences on syntactic complexity, the density of noun phrases, referential cohesion, and many other characteristics. Another important component in Coh-Metrix is the latent semantic analysis (LSA) module that measures the conceptual similarity between sentences, paragraphs, and texts on the basis of world knowledge (Kintsch, 1998, Landauer et al., 1998).

It is well beyond the scope of this chapter to define and discuss the various measures of Coh-Metrix. Instead, we will present an example analysis that scales texts on referential cohesion, causal cohesion, and a few other measures. Consider the four texts that were presented earlier when we discussed how the causal chain of a cylinder lock would be articulated. Table 3 presents a subset of the Coh-Metrix measures for these four texts. The Flesch Kincaid Grade Level and the measure of average sentence length would lead one to the conclusion that Texts 0 and 3 are the easiest texts. They would presumably be assigned to poor readers or readers with low world knowledge. They have shorter sentences, which allegedly are less taxing on working memory. However, as we discussed earlier, this may be a misleading conclusion. Texts 0 and 3 have fewer cohesion signals and relations than do texts 1 and 2. When considering referential
cohesion, texts 1 and 2 are substantially higher than texts 0 and 3. Regarding causal cohesion and the LSA metrics, text 1 is greater than text 0, and text 2 is greater than text 3; the mean scores for texts 1 and 2 combined exceeds texts 0 and 3 combined. Therefore, if we used cohesion as a criterion measure rather than readability, we would recommend assigning texts 1 or 2 rather than texts 0 or 3 to beginning readers or to readers with low word knowledge. There obviously is a clash between the readability formulas and the cohesion metrics of Coh-Metrix. Regarding the readers who are more advanced, with more world knowledge, we would recommend texts 0 or 3, just the opposite of the recommendation of a reading expert who advocates readability formulas.

**INSERT TABLE 3 ABOUT HERE**

We should say a few words about our method of computing referential and causal cohesion. One of the two measures of referential cohesion in Table 3 was computed for adjacent sentences in the text. A pair of adjacent sentences was scored as being referentially linked if (a) there was at least one noun shared by the two sentences (or what some researchers call argument overlap) or (b) the morphological stem of a noun in sentence A overlapped the stem of any content word in sentence B. The “b” alternative allows a referential link between the cam rotated in one sentence and the rotation in another sentence. Our first measure of referential cohesion was simply the proportion of adjacent sentence pairs that had referential links, as defined above. The second measure of referential cohesion was measured in the same way, except that we included sentence pairs that were within two sentences of each other, e.g., sentences 1 and 3 but not sentences 1 and 4.

Causal cohesion was measured as a ratio of (a) the number of words that signal causal cohesion and (b) the number of main verbs that are classified as causal. Examples of words that
signal causal cohesion are conjunctions (e.g., because, so, consequently) and verbs that directly
denote or assert a causal relation (e.g., cause, enable, allow). Verbs were scored as being causal
(e.g., turn, pull, move) if WordNet identified them as being causal.

Global cohesion of a text was measured by Latent Semantic Analysis (LSA). LSA has
recently been proposed as a statistical representation of a large body of world knowledge
(Landauer et al., 1998). LSA uses a statistical method called “singular value decomposition”
(SVD) to reduce a large Word by Document co-occurrence matrix to approximately 100-500
functional dimensions. The Word-by-Document co-occurrence matrix is simply a record of the
number of times word \( W_i \) occurs in document \( D_j \). A document may be defined as a sentence,
paragraph, or section of an article. Each word, sentence, or text ends up being a weighted vector
on the \( K \) dimensions. The “match” (i.e., similarity in meaning, conceptual relatedness) between
two unordered bags of words (single words, sentences, or texts) is computed as a geometric
cosine (or dot product) between the two vectors, with values ranging from 0 to 1. From the
present standpoint, LSA was used to compute the similarity between two sentences, or between
the entire text and a sentence. The first LSA measure was simply the mean cosine value of all
possible pairs of sentences. The second LSA measure was the mean cosine value between each
sentence and the text as a whole.

The measures in Table 3 clearly show that there are trade-offs among some of the
measures of text quality. Texts with short sentences fit within the limited capacity working
memory, but run the risk of sacrificing referential and causal cohesion. Such trade-offs support
the conclusion that there may be no such thing as the perfect text to convey a body of
knowledge. Moreover, an ideal text depends on the characteristics of the reader. Readers with
low reading ability and domain knowledge should be given the texts with higher cohesion and
longer sentences; good readers with high domain knowledge should be assigned texts with lower cohesion to encourage them to actively construct the meaning.

Summary

This chapter has discussed how readers construct situation models while comprehending text. Readers construct different classes of inferences to fill out the situation model. The likelihood of constructing particular classes of inferences is systematic, just like other language modules (such as syntax) because members of a culture are exposed to very similar worlds and have very similar cognitive constraints. However, the alternative models of discourse comprehension make rather different assumptions about what classes of inferences are routinely constructed during comprehension. The construction of the situation model is heavily influenced by linguistic features and cohesion of the text. This chapter identifies many of the linguistic and discourse features that contribute to text cohesion and psychological coherence. We are convinced that the next stage of unraveling the mysteries of situation models, inferences, cohesion, and coherence will require breakthroughs in neuroscience and computational linguistics. Neuroscience will provide more convincing evidence on what cognitive representations are recruited during these comprehension processes. Computational linguistics will offer more precise analytical detail on text characteristics. A recent computer system we have developed, called Coh-Metrix, incorporates new modules and metrics in computational linguistics that allow us to analyze text on over 50 measures of cohesion and 200 measures of language. The question remains whether our metrics of cohesion in Coh-Metrix have any validity from the standpoint of psychological models and brain mechanisms. Both the discoveries and the devil lie in the details.
References

Abney, S. (1997). *The SCOL Manual (version 0.1b)*. Unpublished manuscript, Tuebingen, Germany (www.sfs.nphil.uni-tuebingen.de/~abney/)


Griesel, C., Friese, U. & Schmalhofer, F. (Submitted) What are the differences in the cognitive representations of predictive and bridging inferences?


Author Note
This research was supported by the Institute for Education Sciences (IES R3056020018-02), the National Science Foundation (SBR 9720314, REC-0089271, REC 0106965, REC 0126265, ITR 0325428, ) and the DoD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DoD, ONR, or NSF. We would like to thank members of the Tutoring Research Group (TRG) at the University of Memphis (visit http://www.autotutor.org) and the Cognitive Studies and Educational Practice lab at the University of Memphis (http://csep.psysc.memphis.edu). Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230.
Table 1

*Landscape of Inferences*

<table>
<thead>
<tr>
<th>TYPE OF INFERENCE</th>
<th>BRIEF DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Referential</td>
<td>A word or phrase refers to previous element or constituent in the text.</td>
</tr>
<tr>
<td>2. Case structure role assignment</td>
<td>An explicit noun-phrase is assigned to a particular case structure rule, e.g., agent, recipient, object, location, time.</td>
</tr>
<tr>
<td>3. Instantiation of a noun category</td>
<td>The inference is a subcategory or exemplar that instantiates an explicit noun, e.g., <em>president</em> is inferentially instantiated with <em>George Bush</em>.</td>
</tr>
<tr>
<td>4. Superordinate goal</td>
<td>The inference is a goal that motivates an agent’s intentional action.</td>
</tr>
<tr>
<td>5. Subordinate goal or action</td>
<td>The inference is a goal, plan or action that specifies how an agent’s action is achieved.</td>
</tr>
<tr>
<td>6. Instrument</td>
<td>The inference is an object, part of the body, or resource that is used when an agent executes an intentional action.</td>
</tr>
<tr>
<td>7. Causal antecedent</td>
<td>The inference is on a causal chain that bridges the current explicit action, event, or state to the previous passage context.</td>
</tr>
<tr>
<td>8. Causal Consequence</td>
<td>The inference is on a forecasted causal chain, including physical events, psychological events, and new goals, plans, and actions of agents. Classes 4 and 9 are excluded.</td>
</tr>
<tr>
<td>9. Character emotional reaction</td>
<td>The inference is an emotion experienced by an agent, immediately caused by or in response to an event or action.</td>
</tr>
<tr>
<td>10. Emotion of reader</td>
<td>The inference is an emotion the reader experiences, or is intended to experience, while reading a text.</td>
</tr>
<tr>
<td>11. State</td>
<td>The inference is an ongoing state, from the standpoint of the text, that is not causally related to the story plot. The states include agent’s traits, knowledge, and beliefs; the properties of objects and concepts; and the spatial location of entities.</td>
</tr>
<tr>
<td>12. Themes</td>
<td>This is a main point or moral of the text.</td>
</tr>
<tr>
<td>13. Author’s intent</td>
<td>The inference is the author’s attitude or motive in writing the text.</td>
</tr>
<tr>
<td>TYPE OF SIGNAL OR RELATION</td>
<td>BRIEF DESCRIPTION</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>1. Coreference</td>
<td>Two words or expressions in a text refer to the same person, thing, abstract concept, or idea.</td>
</tr>
<tr>
<td>2. Deixis</td>
<td>References to people, location, and time in a conversation among participants. For example, there are pronouns that refer to people (<em>I, you, we</em>), to location (<em>here, there</em>), and to time (<em>now, then, later</em>).</td>
</tr>
<tr>
<td>3. Given-new cues</td>
<td>The content of a sentence in a text can be segregated into given (old) information and new information. Given information has already been introduced, mentioned, or inferred from the previous text.</td>
</tr>
<tr>
<td>4. Punctuation</td>
<td>In addition, to periods (.), commas (,), colons (:), semi-colons (;), question marks (?), and exclamation points (!), there are quotation marks (“”).</td>
</tr>
<tr>
<td>5. Conjunctive relations</td>
<td>Conjunctive relations are text-connecting relations that normally link adjacent clauses or sentences. The subcategories of conjunctive relations include additive (<em>and, also, moreover</em>), temporal (<em>and then, when, before</em>), causal (<em>because, consequently, as a result</em>), intentional (<em>in order to, by means of</em>), adversative (<em>however, but, although</em>), and logical (<em>therefore, so</em>).</td>
</tr>
<tr>
<td>6. Verb tense and chronology</td>
<td>Events in text often unfold in a chronological order that matches the order of explicit mention. When the order of mention deviations from chronological order (including flashbacks and flash forwards), this is cued by shifts in verb tense and temporal expressions.</td>
</tr>
<tr>
<td>7. Scene changes</td>
<td>A scene is the spatial context that houses the agents in a story or the entities in an expository text. Scene changes are explicitly signaled (<em>meanwhile back at the ranch…</em>).</td>
</tr>
<tr>
<td>8. Topic sentences</td>
<td>The first sentence in a paragraph normally captures the main topic of the paragraph, with subsequent sentences embellishing the topic sentence.</td>
</tr>
<tr>
<td>9. Headers and highlighting</td>
<td>Headers, sub-headers, and highlighted words help organize the text, direct the reader’s attention, and cue discourse genre.</td>
</tr>
<tr>
<td>10. Signals of rhetorical structure</td>
<td>There are distinctive words and phrases associated with major subtypes of genres.</td>
</tr>
</tbody>
</table>
Table 3

*Coh-Metrix Measures for Four Texts on the Cylinder Lock*

<table>
<thead>
<tr>
<th>Coh-Metrix Measures</th>
<th>Four Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Readability Measures</td>
<td></td>
</tr>
<tr>
<td>Average sentence length</td>
<td>6.00</td>
</tr>
<tr>
<td>Average word length</td>
<td>3.70</td>
</tr>
<tr>
<td>Flesch Kincaid Grade Level</td>
<td>.80</td>
</tr>
<tr>
<td>Referential Cohesion</td>
<td></td>
</tr>
<tr>
<td>Adjacent sentences</td>
<td>.75</td>
</tr>
<tr>
<td>2 sentence spans</td>
<td>.55</td>
</tr>
<tr>
<td>Causal Cohesion</td>
<td></td>
</tr>
<tr>
<td>Ratio of causal cohesion particles to causal verbs</td>
<td>.50</td>
</tr>
<tr>
<td>Latent Semantic Analysis</td>
<td></td>
</tr>
<tr>
<td>All possible sentence pairs</td>
<td>.31</td>
</tr>
<tr>
<td>Sentences to text</td>
<td>.49</td>
</tr>
</tbody>
</table>