

Using temporal cohesion to predict temporal coherence in narrative and expository texts

NICHOLAS D. DURAN, PHILIP M. MCCARTHY, ART C. GRAESSER, AND DANIELLE S. MCNAMARA
University of Memphis, Memphis, Tennessee

We investigated the linguistic features of temporal cohesion that distinguish variations in temporal coherence. In an analysis of 150 texts, experts rated temporal coherence on three continuous scale measures designed to capture unique representations of time. Coh-Metrix, a computational tool that assesses textual cohesion, correctly predicted the human ratings with five features of temporal cohesion. The correlations between predicted and actual scores were all statistically significant. In a complementary study, we explored the importance of temporal cohesion in characterizing genre. A discriminant function analysis, using Coh-Metrix temporal indices, successfully distinguished the genres of science, history, and narrative texts. The results suggested that history texts are more similar to narrative texts than to science texts in terms of temporal cohesion.

One important goal in reading comprehension is to build a dynamic mental representation, or situation model, of the state of affairs described in a text (Johnson-Laird, 1983; Kintsch, 1998). One theory that accounts for this process, the event-indexing model (Zwaan, Langston, & Graesser, 1995), posits that readers naturally monitor story events and link these events according to continuities in time, space, causality, intentionality (motivation), and protagonists. Whereas the construction of a situation model is facilitated by readers' verbal skills and general knowledge, the continuities are made explicit by corresponding linguistic features that guide the *when* (time), *where* (space), *who* (protagonists), and *why* (causality and motivations) of event integration (Graesser, Singer, & Trabasso, 1994; Magliano, Zwaan, & Graesser, 1998). Accordingly, linguistic features are critical for creating and maintaining conceptual links between events, thus assisting the reader in forming coherent interpretations of the text (Kintsch, 1998; Magliano & Schleich, 2000).

The purpose of this study is to examine the contribution of temporal linguistic features to evaluations of coherence. We specifically focus on the temporal dimension because of its seemingly ubiquitous presence in organizing language. Time is uniquely represented, via the inflected tense morpheme (e.g., *-ed*, *is*, *has*), in every sentence of the English language (Comrie, 1985). The temporal dimension also depicts unique internal event time frames, such as an event that is complete (*telic*) or ongoing (*atelic*), by incorporating a diverse tense–aspect system (ter Meulen, 1995). Lastly, the occurrence of events at a point in time can be widely established by a large repertoire of adverbial cues such as *before*, *after*, or *then* (Klein, 1994). These temporal features, taken as a whole, constitute the temporal cohesion of a text.

Cohesion can be viewed as the text-based information that explicitly connects constituents, propositions, conceptual themes, and subthemes (Kintsch, 1998). Although this information refers to elements in a text, there is a large body of research consistent with the notion that differences in cohesion correlate with the coherence involved in the mental representation of a text (Beck, McKeown, Sinatra, & Loxterman, 1991; Britton & Gülgöz, 1991; McNamara, Kintsch, Songer, & Kintsch, 1996). As such, a text passage with greater cohesion can assist readers in generating inferences and bridge conceptual gaps, thereby improving comprehension (McKeown, Beck, Sinatra, & Loxterman, 1992; McNamara, 2001; McNamara & Kintsch, 1996). However, although the effects of coreference cohesion indices, such as argument overlap (Kintsch & van Dijk, 1978) and anaphora (Givón, 1992), have been well tested, there has not been the same focus on how temporal features influence coherence, particularly in naturalistic texts.

The lack of research on temporal cohesion may have been the result of computational limitations that have restricted the range of textual inquiry to relatively few passages of manipulated text (Gernsbacher, 1996; Linderholm et al., 2000). However, advances in technology now allow for large corpora to be annotated for the presence of grammatical and lexical features that play an important role in situation model construction. Graesser, McNamara, Louwerse, and Cai (2004) have integrated these text-based linguistic features, as well as other indices of cohesion, readability, and language, into a Web-based software tool called Coh-Metrix (cohmetrix.memphis.edu). One of the benefits of the tool has been its ability to assess textual cohesion, including temporal cohesion. As such, Coh-Metrix offers an ideal approach for addressing

N. D. Duran, nduran@mail.psyc.memphis.edu

our primary research question: Are temporal, text-level features of temporality critical for coherent representations? If so, we expect the profile of indices in Coh-Metrix to predict human evaluations of temporal coherence.

In this study, we tested Coh-Metrix assessments against psychological gold standards of temporal coherence. To compile these gold standards, we developed comprehensive rating scales for collecting human judgments of coherence. Unlike previous studies that have focused on a single criterion for temporal coherence (Scott Rich & Taylor, 2000; Zwaan, Radvansky, Hilliard, & Curiel, 1998), we used three measures motivated by a situation model framework. This framework outlines a process by which readers make use of linguistic features, as well as background knowledge, to link incoming events to their shared event structures in working memory (Zwaan & Radvansky, 1998). It is at the level of the final, integrated event structure that we asked our judges to assess temporal coherence. The judgments were collected over a large corpus of narrative, history, and science texts and applied to a machine learning approach for estimating the extent to which Coh-Metrix mimics psychological evaluations of time.

We fortunately managed to validate the psychological reality of Coh-Metrix on a number of temporal indices. In addition to this primary objective, we were able to conduct a genre identification task that further explored temporal cohesion and coherence. There is broad consensus that genre identification is a skill that readers need to learn in order to process coherence relationships (Bhatia, 1997; Graesser, Olde, & Klettke, 2002). The genre of a text activates particular expectations and strategies that facilitate reading comprehension and assist in the encoding and retrieval of content from episodic long-term memory (van Dijk & Kintsch, 1983). McDaniel, Einstein, Dunay, and Cobb (1986) propose that the most appropriate genre distinction, particularly for differences in comprehension, is between narrative (e.g., fictional literature) and expository (e.g., science, history) texts. Narrative texts encourage readers to process the global and thematic relationships in a passage, most likely because the structure of narrative is easily mapped onto everyday experience (Otero, Leon, & Graesser, 2002). Expository texts, on the other hand, are on topics with which most readers are unfamiliar. The lack of prior knowledge encourages a diligent reader to process details, such as connections between adjacent clauses, low-frequency technical terms, and local information.

For this second study, we evaluated the necessity of temporal cues in identifying narrative and expository texts. Fleischman (1990) argues that temporal features are central to characterizing a narrative norm, because narratives tend to develop in a dynamic temporal order. Conversely, expository texts tend to be bound to static time, as evidenced by a reportative rhetorical structure (Friedman, 1990). Considering the distinctiveness of time in narrative and expository texts, as well as the importance of readers' ability to identify these genres, we simulated genre identification with a simple learning algorithm trained on Coh-Metrix temporal features. By performing this analysis, we addressed the temporal indices that characterize different

genres, with the goal of providing text-level, temporal features that may serve as cues to assist a reader in genre identification.

Temporal Features as Processing Instructions

As mentioned earlier, readers construct a situation model by indexing the shared links between events as they encounter them in a text (Zwaan et al., 1995). With respect to time, the shared information is moderated by three major processing cues: tense, aspect, and adverbial temporal phrases (Zwaan & Radvansky, 1998). These temporal features are regarded as processing instructions on how to integrate incoming information into prior event contexts (Gernsbacher, 1996). As a result, coherence emerges as temporal information is consistently linked throughout a discourse (Trabasso, Suh, & Payton, 1995).

Tense is a temporal feature that usually locates an event's occurrence in the past, present, or future, and tends to be established around a referential point, such as time of utterance. This results in a temporal organization that places events along a continuum, thus affecting the integration of information in working memory (Rinck, Hahnel, & Becker, 2001). Events that occur within a common time frame are more likely to be processed and updated into a unique situation model. For example, Carreiras, Carriedo, Alonso, and Fernandez (1997) manipulated the tense morpheme to allow associations between a character and a job description to be applied in the present (e.g., *Marta now works as an economist*) or separated by a lapse in time (e.g., *In the past Marta worked as an economist*). The recall of the character's occupation was faster and more accurate when the association was depicted in the condition with close temporal proximity. Radvansky, Zwaan, Federico, and Franklin (1998) also manipulated verb tense to create three scenarios in which an agent carried out short-duration activities in a common time frame (e.g., *The grocer was folding a towel/clearing his throat/listening to the radio*) or in multiple time frames (e.g., *The sailor was/is/will be buttoning his shirt/mumbling to himself/daydreaming*). The activities in a common time frame allowed participants to index events in one situation model, whereas the multiple time frames required participants to index events in multiple situation models. A subsequent memory probe confirmed this event organization. Recalling the activities of an agent across multiple time frames was more difficult than recalling activities in a single time frame, supposedly because of interference with competing situation models.

Adverbials such as *in a moment*, *five minutes later*, and *the next day* are other temporal features that can be used to manipulate the temporal proximity of events. As the chronological distance between events increases, the activation of a single situational representation should decrease. Indeed, Zwaan (1996) found that the mental representation of events was adversely affected by adverbials that imposed a greater gap in discourse time. Reading times of critical sentences were slower after a large time shift between narrated events, such as with the two events *John was beaming. An hour later. . .*, than a short time shift, such as with *A moment later*. This finding is

similar to an earlier study by Anderson, Garrod, and Sanford (1983) that demonstrated that an event is processed faster when it is introduced in the text by an adverbial that indicates a time frame within the typical duration of that event. For example, if the event is “At the cinema” and the first introductory sentence is *John walked into the theater*, participants would read a subsequent sentence that occurs in the typical time frame of watching a movie, such as *Two hours later, the credits were rolling*, rather than a sentence that does not occur in the typical time frame of watching a movie, such as *Two days later, the credits were rolling*. Taken together, these two studies (Anderson et al., 1983; Zwaan, 1996) indicate that adverbials are monitored by readers and used to construct a situation model composed of temporally contiguous and plausible event units.

Although tense and adverbials are important for relating events to a particular point in time, it is also necessary to convey the dynamics of the point itself. Hence, the internal duration of an event is conveyed with the use of grammatical aspect (Klein, 1994). For example, perfective aspect, such as the word *Painted* in the sentence *Mary has painted the house*, generally identifies the action of painting as having been completed with an effect remaining in the present (i.e., the house, now, is painted), whereas the use of the present participle aspect, such as *painting* in the sentence *Sam is painting the house*, generally distinguishes the action of painting as ongoing. Magliano and Schleich (2000) emphasized the importance of aspect as a cue for maintaining information in working memory. Perfective events are processed as completed and decay faster in working memory than do ongoing present participle events. Aspect induces readers to tag information in their current mental representation, since it might be relevant for connecting subsequent discourse.

Measures of Temporal Coherence

The temporal features that contribute to constructing and organizing a situation model also influenced the development of our three measures of temporal coherence. In the present study, each judgment measure was implemented on three separate scales that could be rated for degree of importance. The ratings required participants to explicitly monitor the temporal relationships that naturally occur in text. As in previous studies, participants were instructed to attend to one situational-model dimension while reading for comprehension (Therriault, Rinck, & Zwaan, 2006), and to assess their perceived coherence on a rating scale (Scott Rich & Taylor, 2000). However, where Scott Rich and Taylor evaluated the coherence after reading a sentence, our measures were designed so that temporal coherence would be evaluated after processing the entire text.

The temporal coherence judgments made by participants were qualified on three scales that cover a wide range of possible interpretations. These include (1) how well the events in the text are ordered on a dynamic time line, (2) how well the events in the text, if ordered on a dynamic time line, correspond to an iconic sequence, and (3) how well the events in the text, if ordered on a dynamic time line, are explicitly organized with temporal mark-

ers. In the following sections, we explain and describe in greater detail each of these ratings.

Judgment Measure 1: Dynamic versus static time.

Events are temporally ordered on either a dynamic or static time line. In a dynamic temporal sequencing, a continuous relationship exists between events as they occur at changing points in time (Gennari, 2004). For example, in the subordinate construction *The news reported that Bush visited Africa*, two events are described so that the main event, *The news reported*, occurs after, and is predicated upon, the subordinate event of the visit. It is important to note that dynamic time is not confined to subordinate constructions, but includes all event pairs that are temporally contingent on one another. Dynamic time is characterized by Talmy (2000) as “temporal flow,” and is interpreted as such by adaptive changes in verb aspect and verb tense.

The following passage exemplifies the use of temporal grammatical features in a dynamic time line:

They had boarded a tramp steamer in Vladivostok, and the crossing to Japan was crowded and noisy. Igor would be sick the entire trip. The boat was lurching with unrelenting ferocity.

The internal duration of each event can be represented as either having a discrete beginning and end, or as being continuous and open ended (ter Meulen, 1995). For example, *They had boarded a tramp steamer* is a bounded event characterized by the use of perfective aspect. On the other hand, the use of the past progressive in *The boat was lurching with unrelenting ferocity* suggests an ongoing, unbounded event. In both cases, the use of linguistic cues allows for the interpretation of events that are happening or were happening in different time frames. Furthermore, these two events (i.e., boarding a boat and the lurching of the boat) are temporally dependent in the overall temporal structure because of the consistent use of the past tense (Gennari, 2004).

Contrast the above description with the following static time line passage:

Hereditary information is contained in genes, located in the chromosomes of each cell. Each gene carries a single unit of information. An inherited trait can be determined by one or by many genes, and a single gene can influence more than one trait.

Events in this passage are bound to a generic time that is descriptive or reportative. Although the local order can be inferred for causal relationships, the global structure of events is not uniquely sequenced for narrative relevance (Fleischman, 1990). The order of descriptions, as well as the internal event representations, is static or timeless. The linguistic characterization of a static time line is typified by the present tense with no salient aspectual cues (Friedman, 1990).

Additionally, prototypical, everyday experience is aligned with dynamic time (Freyd, 1987). Events are experienced naturally as a narrative. The parallel between the dynamic event sequencing in a text and our experience in the world has the potential to increase the temporal coherence of a situational representation. This text-to-experience

mapping is moderated by linguistic temporal features (i.e., tense and aspect) that human raters monitor in making judgments of coherence. Within a text there may be varying degrees of dynamic or static time. We hypothesized that human judges would agree that a text with a greater relative proportion of dynamic time was high in temporal coherence and that, conversely, a text with a greater relative proportion of static time was low in temporal coherence.

Judgment Measure 2: Iconicity assumption. Another attribute of dynamic time is that events depicted in a text may be sequenced in true chronological order. In fact, comprehenders assume that events described in a text correspond to chronological time because it maps onto their own experience. This is referred to as the *iconicity assumption* (Fleischman, 1990; Hopper, 1979): The completion of one event is marked by the inception of the next event.

Characteristics of the iconicity assumption are evident in the following passage:

Simon had pulled off the jacket, and was turning a sleeve of it inside out; Martina then seized the jacket and it burst its seams. She snatched it up, threw it over her head, and went to the door.

The grammatical temporal features such as verb aspect and tense allow event progression to be aligned with true chronological order. For example, the past perfective, in the initial phrase *Simon had pulled off the jacket*, completes one event and allows subsequent events—turning a sleeve, seizing the jacket—to progress naturally.

The baseline of iconic event presentation can be contrasted with what Talmy (2000) describes as a “helter-skelter jumble.” This jumble includes events that overlap in occurrence, follow disjointed flashback and flashforward sequences, or contain unexplained durations (Ohtsuka & Brewer, 1992). The fact that these event orders are considered to be out of sync implies that there is a baseline of sequential order that can be violated.

In the following passage, event presentation violates iconicity by way of a disjointed sequence:

By the time John noticed the doorbell, it had already rung four times. As usual, he had been listening to loud music on his radio. He turned the radio down and stood up to answer the door.

The event presented in the second sentence—listening to music—occurs before the event in the first sentence—noticing the doorbell ringing. Tense and aspect are important here for cuing a reader that true chronological order is not being followed. For example, the use of the past perfect continuous *had been listening* in the second sentence *As usual, he had been listening to loud music on his radio* suggests that the action of listening to music was already underway by the time another action, the doorbell ringing, began.

When an iconicity violation occurs, there are moderate consequences for text processing. Ohtsuka and Brewer (1992), for example, manipulated event order in a story passage and found that greater deviations from iconic representation resulted in participants committing more

errors while recalling the overall temporal structure. In a similar study, Mandler (1986) also reported that events are read slower when there are chronological discontinuities between them. Although both studies suggest that the construction of a situation model is facilitated when one event can be integrated into the most recent event, there is also facilitation when the events are described as being contiguous (Anderson et al., 1983; Dowty, 1986). Zwaan (1996) referred to this as the *strong iconicity assumption*. Again, the role of temporal linguistic features is relevant for influencing coherence. Duration between events can easily be manipulated by such temporal expressions as *in a moment* or *the next day*.

For the present article, the increased performance for chronological and contiguous event order suggests that judgments of temporal coherence will be influenced likewise. We hypothesized, therefore, that texts with event sequences that follow a chronological order would be judged as higher in temporal coherence than would texts expressed with an out-of-sync chronological order.

Judgment Measure 3: Temporal marker salience. The construal of events on a dynamic time line with true chronological order is the exception, not the norm, in a great deal of published popular fiction (Riessman, 1993). Authors often exercise their artistic license by juxtaposing event order to emphasize different aspects of meaning, defying expectations for dramatic memorableness (Bal, 1985). Although true chronology is the psychological default, comprehenders can still construct situation models along different time line representations. However, when iconicity is overridden, comprehenders often use temporal markers to anchor events in an underlying chronological order. These markers include a variety of grammatical and lexical items, such as adverbials and temporal expressions.

The following passage, for example, is not in true chronological order, but contains explicit cues for placing events on an interpretable temporal time line:

For many years, the Xyagen people struggled with the Galactic government for the right to govern themselves. They dreamed of a time when they could control their lives again. This dream came true on April 1, 2089, just after 12 midnight. The Xyagen celebrated with fireworks in the frontier town called Iqaluit, the very place where the first battle had been fought in 2083.

The above story of Xyagen contains many relevant temporal markers for anchoring events on a mental time line. These include a period of time, an adverbial phrase (e.g., *for many years*), absolute dates (*April 1, 2089; 2083*), an exact time phrase (*12 midnight*), and relative temporal adverbs (e.g., *after*). The sum effect of these cues is to improve temporal coherence (Gernsbacher, 1996); therefore, we expect human judges to rate a passage such as this higher for coherence than they would rate a text with fewer explicit temporal markers.

To reiterate, our three judgment measures of temporal coherence can be summarized on the following scales:

Judgment Measure 1, dynamic versus static time: How well are the events ordered on a dynamic time line?

Judgment Measure 2, iconicity assumption: How well are the events, if ordered on a dynamic time line, matched to an iconic sequence?

Judgment Measure 3, temporal marker salience: How well are the events, if ordered on a dynamic time line, explicitly organized with temporal markers?

We have made the case that text-level, temporal cohesion cues are necessary for guiding evaluations of coherence. However, the question remains: Just how effective are these cues in predicting human judgments, particularly when judges are not explicitly made aware of temporal cohesion? We addressed this question in the first of our two studies using Coh-Metrix.

Coh-Metrix

Coh-Metrix harnesses sophisticated developments in computational linguistics and discourse processing, featuring advanced syntactic parsers (Charniak, 1997; Sekine & Grishman, 1995), part-of-speech taggers (Brill, 1995), latent semantic analysis (LSA; Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2007), and computational modules at other levels of language and discourse analysis. Word and conceptual indices are derived from the WordNet lexical database (Miller, 1990), the MRC database (Coltheart, 1981), and other lexicons available electronically. Coh-Metrix also has conventional metrics of readability, such as Flesch-Kincaid grade level (Klare, 1963). In total, Coh-Metrix has over 400 indices of language, text, and readability (Graesser et al., 2004).

Coh-Metrix has been used in dozens of research projects that span text analyses, models of discourse processing, and learning assessment. For example, much of the research by McNamara and colleagues has examined the cohesion of textbooks and the resulting benefit for readers who vary in knowledge and skill (Best, Rowe, Ozuru, & McNamara, 2005; McNamara et al., 1996). In one study, Ozuru, Best, and McNamara (2004) enhanced low-cohesion texts by manipulating surface-level indicators (e.g., replacing pronouns with noun phrases, adding sentence connectives, and repeating key concepts) and used the Coh-Metrix modules that tap these different types of cohesion (e.g., referential overlap, LSA, word frequency) as a manipulation check to ensure that differences between high and low cohesion could be attributed to the target dimensions of cohesion rather than to third variables.

Coh-Metrix can also distinguish pragmatic characteristics implicit in text content, such as authorship and channel of communication. Louwerse, McCarthy, McNamara, and Graesser (2004), for example, investigated the differences in a large corpus of texts, varying in 23 registers (e.g., mystery fiction, face-to-face conversations), with 241 Coh-Metrix linguistic features. With a principal component analysis (PCA), they demonstrated that the distinction between written versus spoken texts accounted for the most unique variance. McCarthy, Lewis, Dufty, and McNamara (2006) also used Coh-Metrix to distinguish texts of different 20th-century authors. They did so even though the individual style of each author was shown to significantly change as the author's style emerged. This

successful application of Coh-Metrix in evaluating cohesion and in differentiating among text types allowed us to proceed confidently with the present analyses.

Temporal Indices

Coh-Metrix temporal indices were used in our two experiments to investigate temporal coherence evaluations and genre identification. In total, we inspected nine temporal indices, six of which were available on the current online version of Coh-Metrix, whereas another three were developed for this study. The method of calculation for the current indices is via a density score that measures the incidence of a particular category per 1,000 words. Such indices have a global scope, in the sense that a measure applies to a text as a whole (Graesser, McNamara, & Louwerse, 2003).

The six indices currently available on the online version of Coh-Metrix can be divided into three categories based on grammatical function: parts of speech, connectives, and ambiguous elements. The temporal part-of-speech indices include incidence of past participles (e.g., *awoken, begun, seen*), incidence of past tense (e.g., *awoke, began, saw*), and incidence of present tense (e.g., *look, move, talk*). The connective indices include incidence of positive temporal connectives (e.g., *before, then, later*) and incidence of negative temporals (e.g., *until*). The ambiguous-elements score includes temporal adverbial phrases, which consist of nonspecific linguistic features (e.g., *at this time, sooner or later*).

As mentioned above, we developed three additional cohesion indices to be incorporated into the Coh-Metrix tool. These additional indices broaden the scope of accounting for the various temporal relations in a text by capturing temporal words that have a high probability of being embedded in temporal expressions (Wiebe, O'Hara, Öhrström-Sandgren, & McKeever, 1998). Our first index is a composite of all words, including specifiers (e.g., *next, following*), deictics (e.g., *yesterday, now*), absolutes (e.g., *1997, Monday*), time of day (e.g., *12:00 AM, noon*), and time periods (e.g., *summer, week*). The remaining indices are derivations of these explicit elements. The first includes a score that combines specifiers and deictics, whereas the second combines absolutes, time of day, and time periods. The method of calculation is a ratio score that takes the instances of a category divided by all words in a text. These additional three indices are necessary to understand the relative importance of text features that explicitly place events on a time line.

EXPERIMENT 1

Experimental Design

The criteria describing our temporal coherence measures were presented to three experts working on discourse processing at the Cognitive Science Educational Practice (CSEP) lab at the University of Memphis. The experts independently assessed 150 narrative and expository texts to establish the gold standards of temporal coherence. The experts were instructed to rate each text on the basis of a general situation model formed while reading the text.

After the ratings were collected, the Coh-Metrix temporal profiles were used in a logistic regression equation to predict the human evaluations.

Corpus Selection

The 150 texts compiled for this study were selected to cover a range of topics and grade levels. We used an automated approach to randomly extract 400-word, paragraph-to-paragraph slices from 27 published textbooks provided by the MetaMetrics electronic book repository. The candidate textbooks covered three prominent genres (science, history, and narrative) and two grade levels (7th–9th grade, or junior high, and 10th–12th grade, or high school). A major problem with selecting naturalistic texts in an automated, random fashion is that the text samples are removed from their overall context. It was necessary for human raters to evaluate each text for “self-contained” topic continuity and coherence. Texts that did not meet these criteria were discarded from further processing. Texts that contained typographical and content errors, as well as poorly formatted sentence and paragraph breaks, were also discarded, for purposes of clarity.

After we cleansed the corpus by removing unacceptable texts, we selected a final set of 150 texts to maximize the uniform representation of genre, grade level, and authorship. Within each of the three representative genres, we selected 25 texts from the junior high grades and 25 texts from the high school grades. There was an additional stipulation that we choose samples from at least three unique textbooks for each 25-text set. This last criterion allowed a broad distribution across authorship styles, thus removing any confounds due to the idiosyncratic writing styles of any one author (see Table 1).

Interrater Reliability

The human measurements of temporal coherence were assessed using a Likert-type scale ranging from 1 (minimum) to 6 (maximum). A bivariate Pearson correlation for each question was conducted between all possible pairs of raters’ responses. Additionally, agreement was analyzed using Cohen’s weighted kappa statistic (Cohen, 1988). This test is beneficial in compensating for the disagreement that is more likely to occur with a continuous numerical scale. If any two raters were below the good threshold (kappa < 0.6) established by Landis and Koch (1977), and/or correlations were not significant at the $p < .05$ level, ratings were then reexamined and scores were reviewed by the three raters until an acceptable level of agreement was achieved (see Table 2).

The initial assessment of interrater reliability for the human judgment measure of *dynamic versus static time*

Table 2
Interrater Reliability for the Judgment Measure of Dynamic Versus Static Time, Iconicity Assumption, and Temporal Marker Salience

		Kappa	Pearson <i>r</i>
Comparison of Dynamic Versus Static Time			
Rater 1	Rater 2	.736	.754*
Rater 1	Rater 3	.699	.700*
Rater 2	Rater 3	.758	.783*
Comparison of Iconicity Assumption			
Rater 1	Rater 2	.412	.363*
Rater 1	Rater 3	.321	.342*
Rater 2	Rater 3	.439	.383*
Comparison of Temporal Marker Salience			
Rater 1	Rater 2	.744	.674*
Rater 1	Rater 3	.498	.541*
Rater 2	Rater 3	.512	.588*

* $p < .001$.

indicated that ratings consistently ranged from good to very good agreement. Judgments between the three possible pairs of raters were all significantly correlated, with all kappa scores above the 0.6 threshold. The results for the human judgment measure of *iconicity assumption*, however, suggested that raters were not as certain in interpreting texts for chronological order. Though these iconicity results were lower than the previous *dynamic versus static time* measure, they remained significantly correlated. The human judgment measure of *temporal marker salience* received ratings that ranged from good to very good kappa agreement, with one of the kappa scores above the 0.6 threshold and the other two slightly below. Despite the lower kappa agreement, correlations between raters were significant.

The average score of all three raters per text was taken as the final gold standard rating of the 150 texts. After discussions to correct for discrepancies, reevaluated scores resulted in significant correlations and kappa scores that were all in good to excellent agreement (above the 0.6 threshold). The final scores constitute the empirically established gold standards for analysis in the linear regression.

Cross-Validation

We applied a cross-validation technique to provide an objective analysis of our data (Witten & Frank, 2005). As such, we used a training set of 100 texts (randomly selected from the 150 texts) to build a prediction equation that estimated the error rate in the remaining 50-text test set. Each text in the training set was represented by a Coh-Metrix temporal profile of linguistic values. The final prediction equation for each of the temporal coherence measures produced ratings that could be correlated against the true human ratings in the test set. High correlations using a cross-validation technique indicate that the prediction equation is accurate and generalizable.

From the size of the present training set, we estimated that five Coh-Metrix indices would be the maximum number of variables available before problems with overfitting occurred. *Overfitting* refers to capturing unwanted noise in the training set and exaggerating the error rate when transferred to a test set.

Table 1
Distribution of Unique Textbooks and Text Segments for Junior High (Grades 7–9) and High School (Grades 10–12)

	History		Narrative		Science	
	Jr. High	High	Jr. High	High	Jr. High	High
Unique books	5	5	5	4	5	3
Total texts	25	25	25	25	25	25

We selected variables from the three categories of grammatical function—parts of speech, connectives, and ambiguous elements—as well as representatives from the new indices of explicit temporal elements. The variable with the highest correlation to the gold standard human measures was selected from each group as the predictor variable. Other variables were added, provided they were not collinear (i.e., $r > .7$) with the variables that have the highest correlation (Hair, Anderson, Tatham, & Black, 1998). As a result, the five following predictor variables were included in the multiple regression equation: incidence of temporal expression words, incidence of positive temporal connectives, incidence of past tense, incidence of present tense, and temporal adverbial phrases of nonspecific linguistic features.

Results and Discussion

In order to build a prediction equation, a series of forward-entry linear multiple regression analyses was conducted with each of the three temporal judgment measures—*dynamic versus static time*, *iconicity*, and *temporal marker salience*—as dependent variables. A significant overall model emerged for each measure: *dynamic versus static time* [$F(5,99) = 40.51, p < .001$], *iconicity measurement* [$F(5,99) = 14.18, p < .001$], and *temporal marker salience* [$F(5,99) = 23.34, p < .001$]. The (adjusted) variance accounted for in each model is 64.3%, 40%, and 52.4%, respectively.

The multiple regression analyses produced a set of unstandardized beta weights based on the five Coh-Metrix predictor variables (see Table 3).¹ After removing the beta weights that were not significant, the remaining weights were multiplied by their corresponding Coh-Metrix scores in the 50-text test set and added together with the constant to create prediction scores for each judgment measure. These scores were then correlated with the actual human scores to determine the degree to which Coh-Metrix temporal indices mirrored human performance. The correlations comparing the predicted scores and actual scores were all highly significant. For the judgment measure of *dynamic versus static time*, the correlation between predicted scores and actual scores was $r(48) = .850, p < .001$. For the judgment measure of *iconicity assumption*, the correlation between predicted and actual scores was $r(48) = .495, p < .001$. For the judgment measure of *temporal marker salience*, the correlation between predicted

and actual scores was $r(48) = .766, p < .001$. Overall, these correlations provide evidence that distinctions made by human raters can be successfully identified by the Coh-Metrix cohesion indices.

Insofar as the temporal judgment measures are proxies of temporal coherence, the weights in Table 3 also provide insight into the relative contributions of each temporal feature on ratings of temporal coherence. For the dynamic versus static time judgment, higher ratings correspond to an increase in the incidence of temporal expression words and the incidence of past tense. Because dynamic time usually depicts actions and progressions that happen across multiple time lines, the past tense, coupled with temporal words that emphasize event occurrences, is an appropriate cue for guiding judgments (Talmy, 2000). In comparison, the iconicity assumption judgment was not significantly influenced by temporal expression words, since the depiction of iconicity might instead be realized in a more global, top-down fashion. This is evidenced by the increase in ratings that correspond to a greater incidence of past tense. However, the additional influence of temporal adverbial phrases of nonspecific linguistic features (e.g., *sooner or later, in a moment*) is harder to explain. We speculate that a highly iconic temporal representation of events is perceived as dull, as suggested by Bal (1985), and, in an attempt to interject the story with novelty, the writer relies on nonspecific adverbial phrases typical of narrative stylistics (Riessman, 1993). Lastly, for the temporal marker salience judgment measure, higher ratings correspond to a higher incidence of temporal expression words, a decrease in the incidence of present tense, and an increase in the incidence of past tense. Although the high incidence of temporal expression words was not surprising, since these words are the most explicit cues for grounding events on a mental time line, the increasing past tense and decreasing present tense appear to facilitate using temporal markers for making coherence judgments.

Overall, the cohesion indices that emerged from this analysis as being most indicative of the gold standard human judgments were incidence scores for part of speech (except for incidence of the past participle) and nonspecific linguistic features, as well as ratio scores for temporal expression words. The two connective indices, incidence of positive temporals and incidence of negative temporals, were not significantly predictive.

Table 3
Beta Weights of Temporal Coherence Measurements Regressed on Coh-Metrix Indices for the 100-Text Training Set

Coh-Metrix Temporal Indices	Dynamic vs. Static Time	Iconicity Assumption	Temporal Marker Salience
Temporal expression words	.295**	.024	.447**
Adverbial phrases/nonspecific	.141	.303**	.061
Incidence of present tense	-.163	-.100	-.324**
Incidence of past tense	.579**	.458**	.254*
Positive temporal connectives	-.047	-.143	.008
Constant	2.143	2.297	2.748

Note—For ease of comparison, standardized beta weights are presented in the table. However, beta weights were entered into the prediction equation reported in this study. * $p < .05$. ** $p < .001$.

EXPERIMENT 2

In this study, we were interested in cataloging characteristics of text genre with indices of temporal cohesion. Lightman, McCarthy, Dufty, and McNamara (2007) have successfully shown that cohesion rates in academic textbooks are consistent with the grade level of difficulty but that they vary according to genre (e.g., science and history). We expanded this approach by examining the temporal differences in our comprehensive corpus of science, history, and narrative texts. We were particularly interested in issues concerning the text-level features that characterize different genres. These differences are believed to contribute to text comprehension by cuing readers’ natural processing strategies, which are most effective for integrating the event structure in narrative or expository texts (McDaniel et al., 1986). However, it remains unspecified which temporal features are suitable candidates for distinguishing genre. It is important to identify these linguistic features, because they will end up being salient cues to a reader in strategy activation. On the basis of research by Fleischman (1990) and Friedman (1990), we hypothesized that temporal information is necessary for differentiating narratives and expository text. Using the Coh-Metrix temporal indices established in our previous studies, we further explored the contribution of temporality in text.

Experimental Design

A discriminate-function analysis was conducted to examine whether genre is distinguishable by Coh-Metrix indices of temporal cohesion. To this end, the genre corpus of 150 texts (50 narrative, 50 science, and 50 history) was split into two random groups to create a training set and a test set. The purpose of the training set was to allow the statistical model to “learn” which of the seven temporal indices would most effectively differentiate the three genre categories. The relevant indices were then applied to the test set for classifying appropriate genre. A total of 100 texts, distributed equally across genre, comprised the training set. The remaining 50 texts were used to cross-validate the predicted genre classification.

Relevant Predictors

A MANOVA was conducted with the seven temporal indices to determine which were the most significant in differentiating genre. This served the purpose of narrowing the predictors to a smaller number to avoid problems with overfitting unwanted noise. We assumed five vari-

ables would be most appropriate on the basis of an optimal ratio (20:1) of variables to data points (100 texts), since this ratio is standard in many machine learning protocols (Witten & Frank, 2005). Indices that did not exceed a statistical significance value of $p < .05$ were discarded. The correlations between all indices were also at $r < .70$, satisfying an assumption required for discriminate analysis that predictors not be highly correlated. Consequently, two indices were not included as predictors in the discriminate analysis: the incidence of past participles and the incidence of negative temporal connectives (see Table 4).

Post Hoc Analysis of Relevant Predictors

To more closely assess where differences lay between genres, we also conducted a post hoc Bonferroni analysis on significant indices. This analysis was beneficial in highlighting the degree to which genres differ and the direction of those differences (see Table 5).

The most obvious differences were between narrative and science texts, followed by history and science texts, followed by the differences between history and narrative texts. To further understand the temporal composition of each text, the statistically significant temporal features for each genre were examined.

For temporal adverbial phrases of nonspecific linguistic features (e.g., *in a moment, sooner or later*), there is a much lower incidence for science texts compared with narrative and history texts, although a higher incidence occurs for narrative than for history texts. These findings suggest that nonspecific features are stylistic markers of narration, since science and history texts are more direct in conveying information. The incidence of present tense was greater in science texts than in both history and narrative texts, with no difference between history and narrative. Conversely, for incidence of past tense, there is a stark contrast between narrative texts and science texts, with narratives predominantly written in the past tense. This difference in past tense is also true of narratives compared with history texts. However, history texts had a much higher incidence score of past tense than did the science texts, which had the advantage in present tense. Finally, for positive temporal connectives (e.g., *before, then, later*), narratives had a higher incidence than did science and history texts, with no difference between history and science.

Results and Discussion

Using the five statistically significant temporal predictors, we tested the accuracy of the discriminate function analysis to predict the correct genre membership (17 nar-

Table 4
Results of MANOVA for the Coh-Metrix Temporal Indices As a Function of Genre

Coh-Metrix Temporal Indices	History		Narrative		Science		F	p
	M	SD	M	SD	M	SD		
Temporal expression words	36.74	15.91	29.99	13.93	29.05	16.52	3.37	.04
Adverbial phrases/nonspecific	7.86	5.80	10.53	5.37	5.39	3.92	12.71	0
Incidence of present tense	7.22	11.10	7.55	12.00	30.66	12.83	62.75	0
Incidence of past tense	56.54	22.45	72.17	23.12	5.93	9.81	158.00	0
Positive temporal connectives	7.32	5.30	11.65	5.87	6.85	5.21	11.69	0

Table 5
Bonferroni Post Hoc Analysis Showing
Direction of Differences Between Genres

Coh-Metrix Temporal Indices	History vs. Narrative	History vs. Science	Narrative vs. Science
Temporal expression words	6.748	7.685	0.936
Temporal nonspecific adverbial phrases	-2.672*	2.467*	5.138**
Incidence of present tense	-0.333	-23.445**	-23.116**
Incidence of past tense	-15.620**	50.620**	66.240**
Positive temporal connectives	-4.330**	0.466	4.796**

* $p < .05$. ** $p < .001$.

rative texts, 17 history texts, and 16 science texts). The accuracy of the trained model on the test set is displayed with the actual genre classification against the predicted genre classification (see Table 6). The diagonals represent the frequency with which the genre was correctly identified. The off-diagonals represent the number of incorrect identifications of genre. Overall, 68% of texts were correctly identified in the test set, with 88% of the history texts, 47% of the narratives, and 69% of the science texts accounted for.

The recall, precision, and *F*-measure scores (Jurafsky & Martin, 2000) for each genre also demonstrate the accuracy of the model (see Table 6). The *F*-measure score (a composite score of precision and recall) indicates that there is greater accuracy in identifying science texts (.79) than there is in either narrative (.59) or history (.67) texts. Additionally, the precision scores (correct predictions divided by the number of correct plus incorrect predictions) indicate that there is more of a bias for the Coh-Metrix indices to label a narrative text as a history text (8 false alarms) than to label it as a science text (1 false alarm). This false-alarm rate for narratives would explain the relatively low correct classification of 47% reported above. There is also more of a bias for the Coh-Metrix indices to label a history text as a narrative text (2 false alarms) than to label it as a science text (0 false alarms).

The results lend credence to the hypothesis that temporal text-based features of situation model construction are able to differentiate between genres, and that each genre has a characteristic temporal composition. However, the results of the post hoc Bonferroni and accuracy scores suggest that history and narrative texts have a more similar temporal composition than do science texts. This finding is interesting, in that expository texts are considered to be more difficult to remember than narrative texts (Graesser et al., 2002). Although more research is needed, tempo-

ral dynamics may contribute to a unique situation model construction for science texts, in comparison with narrative or history texts. Furthermore, the overlap in temporal matching between history and narrative texts may require a reassessment of the label traditionally assigned to history texts (Beck et al., 1991).

GENERAL DISCUSSION

In this article, we demonstrated that Coh-Metrix, an automated tool for assessing text cohesion and readability, could generate profiles of temporal cohesion corresponding to human interpretations of temporal coherence. This conclusion is drawn from the successful performance of a linear regression algorithm that predicted gold standards of temporal coherence for 150 naturalistic texts. The predictors supplied to the model were a subset of five temporal cohesion features generated by Coh-Metrix: incidence of temporal expression words, incidence of positive temporal connectives, incidence of past tense, incidence of present tense, and temporal adverbial phrases of nonspecific linguistic features. Collectively, all but one of the predictors—the incidence of positive connectives—provided evidence for a link between profiles of temporal cohesion and temporal coherence. These results contribute to a large body of research that posits that reading comprehension is an interaction between the text (i.e., the text base) and a reader's general knowledge (i.e., situation model). Moreover, this study underscores the importance of temporal features as an implicit and viable cue for facilitating the coherent interpretations of a text.

This study also demonstrated that Coh-Metrix temporal features could distinguish the temporal structure of different genres of texts. A discriminant function analysis with narrative, history, and science texts provided evidence that all genres contained a significant number of distinct features. This distinctiveness held true even though science

Table 6
Classifications of Texts Into Genre Categories As a Function of
Actual Genre and Predicted Genre With Recall, Precision, and
***F*-Measure Scores As a Function of Genre**

Genre	Predicted Genre			Accuracy		
	History	Narrative	Science	Recall	Precision	<i>F</i>
History	15	2	0	.88	.54	.67
Narrative	8	8	1	.47	.80	.59
Science	5	0	11	.69	.92	.79

texts were more successfully classified than were either history or narrative texts. These findings are relevant to reading comprehension, because temporal features specific to a particular genre may contribute to the activation of effective reading strategies. For example, science texts, which are predominantly written in the present tense with few temporal markers, naturally guide a reader to the detailed, local content information that is necessary to make good bridging inferences (Otero et al., 2002). Narratives, on the other hand, which are most often written in the past tense with many temporal markers, may contribute to the more global inferences that good readers often make. Future research, therefore, needs to examine the role of genre-specific temporal features in facilitating different types of inferences.

Along with the findings of our two experiments, the three temporal coherence measures developed for this study also contribute to research in discourse processing. We created three criteria of temporal coherence grounded in 20 years of psychological and linguistic inquiry. However, the common assumption guiding the development of our final three measures (i.e., *dynamic vs. static time*, *iconicity assumption*, and *temporal marker salience*) was straightforward: Events in a text are vicarious experiences. As such, the closer the match between the depicted events and our own experience, the more coherent the text should be. With this heuristic in mind, a temporally coherent text would be assessed on our ratings scales as having a greater proportion of dynamic than static time, an order of mention that corresponds to true chronological order (iconic), and a large repertoire of temporal markers that establish events on an easily retrievable mental time line. These notions of temporal coherence were substantiated on the basis of the agreement of independent ratings provided by expert judges.

A possible limitation with this study, however, is that predicted human scores were successful because humans were evaluating temporal coherence on the same criteria as Coh-Metrix. Although this cannot entirely be ruled out, the expert raters were not made aware that the coherence judgments would be analyzed with the text-level, temporal features in the text. The instructions to the raters were intentionally limited to the three Likert scales that evaluated perceived coherence (as opposed to perceived cohesion) of an entire text.

Future research will focus on the development of even more sophisticated Coh-Metrix temporal indices motivated by a situation model framework. These new indices will capture both the grammatical mode and global structure of temporal cohesion relations. They will be calculated by tracking the repetition of tense and aspect across paragraphs and documents. Repetition is important, because integrating events into a coherent situation model is facilitated by the consistency of temporal features (Rinck et al., 2001). Furthermore, the new indices will acknowledge the role of temporal adverbials in grounding events along a mental time line. An adverbial index will be calculated as the number of adverbials over the consistency score of tense and aspect. In this way, a text with low consistency can be considered temporally coherent as long as there are a sufficient number of adverbials to com-

pensate for shifts in tense and aspect. These new indices will add further discriminatory power to identifying texts at various levels of temporal coherence. They will also be particularly beneficial to our assessment of coherence measures that rely on temporal consistency, such as the iconicity assumption.

A further goal for our temporal investigations is to address the concern that situation model dimensions are studied in isolation, whereas coherence is dependent on the interaction of dimensions (Zwaan & Radvansky, 1998). Temporal relationships, for instance, are also influenced by causal links between events. Because Coh-Metrix also provides a wide variety of causal and intentional indices, we believe that the representations of temporal and causal cohesion can be assessed simultaneously. This may provide insight concerning the corelevance of these dimensions for comprehension of a text.

In conclusion, Coh-Metrix is highlighted as a viable computational tool for exploring processes involved in reading comprehension. The temporal indices, generated automatically, accurately reflected human evaluations of temporal coherence and genre classification. Although this study is certainly a first step in profiling the linguistic features of situational model dimensions, it brings us closer to understanding the complex relationship that exists between cohesion and coherence, a relationship that, more often than not, occurs in the same naturalistic, unmanipulated texts used in the educational curriculum.

AUTHOR NOTE

The research was supported by the Institute for Educational Sciences (IES R305G020018-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES. We thank Stephen Briner and Erin Lightman for their contributions to this study. Correspondence regarding this article should be addressed to N. D. Duran, Department of Psychology, University of Memphis, Memphis, TN 38152 (e-mail: nduran@mail.psy.memphis.edu).

REFERENCES

- ANDERSON, A., GARROD, S. C., & SANFORD, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, **35A**, 427-440.
- BAL, M. (1985). *Narratology: Introduction to the theory of narrative* (C. van Boheemen, Trans.). Toronto: University of Toronto Press.
- BECK, I. L., MCKEOWN, M. G., SINATRA, G. M., & LOXTERMAN, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly*, **26**, 251-276.
- BEST, R. M., ROWE, M., OZURU, Y., & MCNAMARA, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, **25**, 65-83.
- BHATIA, V. (1997). Applied genre analysis and ESP. In T. Miller (Ed.), *Functional approaches to written text: Classroom applications*. Washington, DC: USIA.
- BRILL, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, **21**, 543-565.
- BRITTON, B. K., & GÜLGÖZ, 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.
- CARREIRAS, M., CARRIEDO, N., ALONSO, M. A., & FERNANDEZ, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, **25**, 438-446.

- CHARNIAK, E. (1997, Winter). Statistical techniques for natural language processing. *AI Magazine*, **18**, 33-44.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences*. San Diego: Academic Press.
- COLTHEART, M. (1981). The MRC psycholinguistics database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.
- COMRIE, B. (1985). *Tense*. Cambridge: Cambridge University Press.
- DOWTY, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics & Philosophy*, **9**, 37-61.
- FLEISCHMAN, S. (1990). *Tense and narrativity: From medieval performance to modern fiction*. Austin: University of Texas Press.
- FREYD, J. J. (1987). Dynamic mental representations. *Psychological Review*, **94**, 427-438.
- FRIEDMAN, W. J. (1990). *About time*. Cambridge, MA: MIT Press.
- GENNARI, S. P. (2004). Temporal reference and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 877-890.
- GERNSBACHER, M. A. (1996). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships in the production and comprehension of text* (pp. 3-21). Mahwah, NJ: Erlbaum.
- GIVÓN, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, **30**, 5-55.
- GRAESSER, A. C., MCNAMARA, D. S., & LOUWERSE, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford.
- GRAESSER, A. C., MCNAMARA, D. S., LOUWERSE, M. M., & CAI, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, **36**, 193-202.
- GRAESSER, A. C., OLDE, B., & KLETTKE, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 229-262). Mahwah, NJ: Erlbaum.
- GRAESSER, A. C., SINGER, M., & TRABASSO, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**, 371-395.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., & BLACK, W. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- HOPPER, P. J. (1979). Aspect and foregrounding in discourse. In T. Givón (Ed.), *Discourse and syntax* (Syntax and Semantics, Vol. 12, pp. 213-241). New York: Academic Press.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- JURAFSKY, D., & MARTIN, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*. Upper Saddle River, NJ: Prentice Hall.
- KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: 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. (1963). *The measurement of readability*. Ames, IA: Iowa State University Press.
- KLEIN, W. (1994). *Time in language*. London: Routledge.
- LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.
- LANDAUER, T. K., MCNAMARA, D. S., DENNIS, S., & KINTSCH, W. (Eds.) (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- LANDIS, J. R., & KOCH, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.
- LIGHTMAN, E. J., MCCARTHY, P. M., DUFTY, D. F., & MCNAMARA, D. S. (2007). The structural organization of high school texts. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 20th Annual Florida Artificial Intelligence Research Society International Conference* (pp. 235-240). Menlo Park, CA: AAAI Press.
- LINDERHOLM, T., EVERSON, M. G., VAN DEN BROEK, P., MISCHINSKI, G., CRITTENDEN, A., & SAMUELS, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition & Instruction*, **18**, 525-556.
- LOUWERSE, M. M., MCCARTHY, P. M., MCNAMARA, D. S., & GRAESSER, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 843-848). Mahwah, NJ: Erlbaum.
- MAGLIANO, J. P., & SCHLEICH, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, **29**, 83-112.
- MAGLIANO, J. P., ZWAAN, R. A., & GRAESSER, A. C. (1998). The role of situational continuity in narrative understanding. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 219-245). Mahwah, NJ: Erlbaum.
- MANDLER, J. M. (1986). On the comprehension of temporal order. *Language & Cognitive Processes*, **1**, 309-320.
- MCCARTHY, P. M., LEWIS, G. A., DUFTY, D. F., & MCNAMARA, D. S. (2006). Analyzing writing styles with Coh-Metrix. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference* (pp. 764-769). Melbourne Beach, FL: FLAIRS.
- MCDANIEL, M. A., EINSTEIN, G. O., DUNAY, P. K., & COBB, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory & Language*, **25**, 645-656.
- MCKEOWN, M. G., BECK, I. L., SINATRA, M. G., & LOXTERMAN, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, **27**, 78-93.
- MCNAMARA, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, **55**, 51-62.
- MCNAMARA, D. S., & KINTSCH, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, **22**, 247-288.
- MCNAMARA, D. S., KINTSCH, E., SONGER, B. N., & KINTSCH, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from texts. *Cognition & Instruction*, **14**, 1-43.
- MILLER, G. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, **3**, 235-312.
- OHTSUKA, K., & BREWER, W. F. (1992). Discourse organization in the comprehension of temporal order in narrative texts. *Discourse Processes*, **15**, 317-336.
- OTERO, J., LEON, J. A., & GRAESSER, A. C. (Eds.) (2002). *The psychology of science text comprehension*. Mahwah, NJ: Erlbaum.
- OZURU, Y., BEST, R., & MCNAMARA, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1071-1076). Mahwah, NJ: Erlbaum.
- RADVANSKY, G. A., ZWAAN, R. A., FEDERICO, T., & FRANKLIN, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1224-1237.
- RIESSMAN, C. K. (1993). *Narrative analysis*. Newbury Park, CA: Sage.
- RINCK, M., HAHNEL, A., & BECKER, G. (2001). Using temporal information to construct, update, and retrieve situation models of narratives. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 67-80.
- SCOTT RICH, S., & TAYLOR, H. A. (2000). Not all narrative shifts function equally. *Memory & Cognition*, **28**, 1257-1266.
- SEKINE, S., & GRISHMAN, R. (1995). A corpus-based probabilistic grammar with only two non-terminals. In *Proceedings of the Fourth International Workshop on Parsing Technology* (pp. 216-223). Prague.
- TALMY, L. (2000). *Toward a cognitive semantics: Vol. 1. Concept structuring systems*. Cambridge, MA: MIT Press.
- TER MEULEN, G. B. A. (1995). *Representing time in natural language: The dynamic interpretation of tense and aspect*. Cambridge, MA: MIT Press.
- THERRIault, D. J., RINCK, M., & ZWAAN, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Memory & Cognition*, **34**, 78-89.
- TRABASSO, T., SUH, S., & PAYTON, P. (1995). Explanatory coherence in understanding and talking about events. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 189-214). Amsterdam: John Benjamins.
- VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- WIEBE, J. M., O'HARA, T. P., ÖHRSTRÖM-SANDGREN, T., & MCKEEVER,

- K. J. (1998). An empirical approach to temporal reference resolution. *Journal of Artificial Intelligence Research*, **9**, 247-293.
- WITTEN, I. H., & FRANK, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann.
- ZWAAN, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1196-1207.
- ZWAAN, R. A., LANGSTON, M. C., & GRAESSER, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, **6**, 292-297.
- ZWAAN, R. A., & RADVANSKY, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**, 162-185.
- ZWAAN, R. A., RADVANSKY, G. A., HILLIARD, A. E., & CURIEL, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, **2**, 199-220.

NOTE

1. For ease of comparison, standardized beta weights are presented in Table 3. However, beta weights were entered into the prediction equation reported in this study.

(Manuscript received December 15, 2006;
revision accepted for publication March 13, 2007.)