

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/220017535>

Latent Semantic Analysis for Text-Based Research

Article in Behavior Research Methods · February 1996

DOI: 10.3758/BF03204765

CITATIONS

193

READS

1,892

1 author:



Peter Foltz

Pearson Inc

139 PUBLICATIONS 7,160 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



LSA Theory [View project](#)

ANALYSIS OF SEMANTIC AND CLINICAL DATA

Chaired by Matthew S. McGlone, *Lafayette College*

Latent semantic analysis for text-based research

PETER W. FOLTZ

New Mexico State University, Las Cruces, New Mexico

Latent semantic analysis (LSA) is a statistical model of word usage that permits comparisons of semantic similarity between pieces of textual information. This paper summarizes three experiments that illustrate how LSA may be used in text-based research. Two experiments describe methods for analyzing a subject's essay for determining from what text a subject learned the information and for grading the quality of information cited in the essay. The third experiment describes using LSA to measure the coherence and comprehensibility of texts.

One of the primary goals in text-comprehension research is to understand what factors influence a reader's ability to extract and retain information from textual material. The typical approach in text-comprehension research is to have subjects read textual material and then have them produce some form of summary, such as answering questions or writing an essay. This summary permits the experimenter to determine what information the subject has gained from the text.

To analyze what a subject has learned from a text, the task of the experimenter is to relate what was in the summary to what the subject has read. This permits the subject's representation (cognitive model) of the text to be compared with the representation expressed in the original text. For such an analysis, the experimenter must examine each sentence in the subject's summary and match the information contained in the sentence to the information contained in the texts that were read. Information in the summary that is highly related to information from the texts would indicate that it was likely learned from the text. Nevertheless, matching this information is not easy. It requires scanning through the original texts to locate the information. In addition, since subjects do not write exactly the same words as those that they have read, it is not possible to look for exact matches. Instead, the experimenter must make the match on the basis of the semantic content of the text.

A theoretical approach to studying text comprehension has been to develop cognitive models of the reader's representation of the text (e.g., Kintsch, 1988; van Dijk & Kintsch, 1983). In such a model, semantic information from both the text and the reader's summary are represented as sets of semantic components called propositions. Typically, each clause in a text is represented by a single proposition. In addition, propositions are linked to each other on the basis of a variety of such criteria as if they share arguments or referents. A complete linking of the propositions from a text will represent the structure of that text. Extensive empirical evidence exists which validates the idea of propositions as psychological processing units (see, e.g., Graesser, 1981; Ratcliff & McKoon, 1978).

Performing a propositional analysis on a text provides a set of semantic primitives that describe the information contained in the text. Similarly, performing a propositional analysis on a subject's recall will provide a set of semantic primitives that provide a representation of the subject's memory of the text. This permits an experimenter to compare the semantic content contained in the text with that in the subject's summary. The advantage of making the comparison at the semantic level is that the comparison is not dependent on surface features, such as the choice of words. Although manuals have been developed to aid in propositionalizing texts (e.g., Bovair & Kieras, 1984; Turner & Green, 1978), propositionalizing texts can be time consuming and require a lot of effort. This can limit the size of texts that are analyzed. Indeed, most research in text comprehension has used texts that are under 1,000 words. In addition, fully computerized methods for generating propositions are not currently feasible since they would require the computer to parse accurately the text and interpret the correct meanings of all words.

While automatic propositionalization is not possible, one of the primary advantages of using propositions is

This work has benefited from collaborative research with Tom Landauer, Walter Kintsch, Charles Perfetti, M. Anne Britt, Susan Dumais, Mara Georgi, and Adrienne Lee. Support for this research was funded by grants from Bell Communications Research and by the Office of Educational Research and Improvement (OERI). Correspondence may be addressed to Peter W. Foltz, Department of Psychology, New Mexico State University, Box 30001, Dept. 3452, Las Cruces, NM 88003-8001 (e-mail: pfoltz@crl.nmsu.edu).

that they can be used for making comparisons of semantic similarity between pieces of textual information. This paper describes an approach to performing semantic matching that can be applied to a variety of areas in text comprehension research that typically use modeling with propositions.

LSA: AN AUTOMATIC METHOD FOR TEXT RESEARCH

Latent semantic analysis (LSA) is a statistical model of word usage that permits comparisons of the semantic similarity between pieces of textual information. LSA was originally designed to improve the effectiveness of information-retrieval methods by performing retrieval based on the derived "semantic" content of words in a query as opposed to performing direct word matching. This approach avoids some of the problems of synonymy, in which different words can be used to describe the same semantic concept. A brief overview of LSA will be provided here. More complete descriptions of LSA may be found in Deerwester, Dumais, Furnas, Landauer, & Harshman (1990) and Dumais (1990).

The primary assumption of LSA is that there is some underlying, or "latent," structure in the pattern of word usage across documents, and that statistical techniques can be used to estimate this latent structure. The term *documents* in this case can be thought of as contexts in which words occur and also could be smaller text segments such as individual paragraphs or sentences. Through an analysis of the associations among words and documents, the method produces a representation in which words that are used in similar contexts will be more semantically associated.

To analyze a text, LSA first generates a matrix of occurrences of each word in each document (sentences or paragraphs). LSA then uses singular-value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis. The SVD scaling decomposes the word-by-document matrix into a set of k , typically 100 to 300, orthogonal factors from which the original matrix can be approximated by linear combination. Instead of representing documents and terms directly as vectors of independent words, LSA represents them as continuous values on each of the k orthogonal indexing dimensions derived from the SVD analysis. Since the number of factors or dimensions is much smaller than the number of unique terms, words will not be independent. For example, if two terms are used in similar contexts (documents), they will have similar vectors in the reduced-dimensional LSA representation. One advantage of this approach is that matching can be done between two pieces of textual information, even if they have no words in common. To illustrate this, if the LSA was trained on a large number of documents, including the following two:

1. The U.S.S. *Nashville* arrived in Colon harbor with 42 marines.
2. With the warship in Colon harbor, the Colombian troops withdrew.

The vector for the word *warship* would be similar to that of the word *Nashville* because both words occur in the same context of other words such as *Colon* and *harbor*. Thus, the LSA technique automatically captures deeper associative structure than simple term-term correlations and clusters.

One can interpret the analysis performed by SVD geometrically. The result of the SVD is a k -dimensional vector space containing a vector for each term and each document. The location of term vectors reflects the correlations in their usage across documents. Similarly, the location of document vectors reflects correlations in the terms used in the documents. In this space, the cosine or dot product between vectors corresponds to their estimated semantic similarity. Thus, by determining the vectors of two pieces of textual information, we can determine the semantic similarity between them.

LSA is well suited for applications for researchers in psychology and education who must assess learning from textual material. By performing an automatic analysis of the texts that were read by subjects, the derived semantic space can be used for matching among pieces of textual information much in the same way as a propositional analysis. This paper summarizes results from three experiments to illustrate applications of LSA for text research. The first and second experiments (Foltz, Britt, & Perfetti, 1994), describe methods for analyzing a subject's essay for determining from what text a subject learned the information and for grading how much relevant information is cited in the essay. The third experiment (Foltz, Kintsch, & Landauer, 1993) describes an approach to using LSA to measure the coherence and comprehensibility of texts.

THREE APPLICATIONS OF LSA TO TEXT RESEARCH

Matching Summaries to the Texts Read

When examining a summary written by a subject, it is often important to know where the subject learned the information reflected in the summary. In research on the subject's reasoning after reading multiple documents, it is similarly important to know which documents have the most influence on the subject's recall. Recent studies of learning from history documents have shown that different types of documents have differing amounts of influence on a subject's reasoning and recall (Britt, Rouet, Georgi, & Perfetti, 1994; Perfetti, Britt, Rouet, Georgi, & Mason, 1994). As part of one of the experiments described by Britt et al., 24 college students read 21 texts related to the events leading up to the building of the Panama Canal. The texts included excerpts from textbooks, historians' and participants' accounts, and primary documents such as treaties and telegrams. The total length of text was 6,097 words. After reading the texts, the subjects wrote an essay on "To what extent was the U.S. intervention in Panama justified?" In the original analysis described by Britt et al., the essays were propositionalized and propositions from the essay were matched against

those in the original texts in order to determine which texts showed the greatest influence in the subjects' essays.

Foltz et al. (1994) reanalyzed the essays, using LSA to make predictions about which texts influenced the subjects' essays. The goal was to match individual sentences from the subjects' essays against the sentences in the original texts read by the subjects. Sentences in the essays that were highly semantically similar to those in the original texts would probably indicate the source of the subject's knowledge.

To perform the LSA analysis, the texts were first run through an SVD scaling in order to generate a semantic space on the topic of the Panama Canal. The 21 texts the subjects read (6,097 words), along with paragraphs from eight encyclopedia articles on the Panama Canal (~4,800 words) and excerpts from two books (~17,000 words) were included in the scaling. Because the semantic space derived by LSA is dependent on having many examples of the co-occurrences of words, the addition of these other textual materials helped to provide the LSA analysis with additional examples of words related to the Panama Canal to help define the semantics of the domain. The LSA analysis resulted in a 100-dimensional space made up of 607 text units by 4,829 unique words.

To analyze the essays, the vector for each sentence from each subject's essay was compared with the vectors for each of the sentences from the original texts read in the derived semantic space. For each sentence, the analysis returned a rank-ordered list of the sentences that matched best on the basis of the cosine between the sentences. For example, for an analysis of the sentence from one of the subject's essays: "Only 42 marines were on the U.S.S. *Nashville*," the best two matches would be the following two sentences:

MF.2.1 Nov. 2, 5:30 PM.: U.S.S. *Nashville* arrives in Colon Harbor with 42 marines. (cosine: 0.64)

P1.2.1. To Hubbard, commander of the U.S.S. *Nashville*, from the Secretary of the Navy (Nov. 2, 1903): Maintain free and uninterrupted transit. (cosine: 0.56)

The codes at the beginning of the returned sentences (MF.2.1 and P1.2.1) indicate which document and which sentence within the document was matched, while the cosines indicate the degree of match. As can be seen, the first document (MF) contains much of the same semantic information as expressed in the sentence from the subject's sentence, and it is highly likely that this document was the source of the subject's knowledge expressed in that sentence.

To determine the effectiveness of LSA's predictions about the source of each subject's knowledge, the predictions were compared against predictions made by two human raters. The raters, who were highly familiar with the topic of the Panama Canal and with the 21 texts, independently read through the essays and, for each sentence, they identified which of the 21 texts was the likely source of the information expressed in the sentence. Because sentences in the essays were complex, often expressing multiple pieces of information, the experimenters

were allowed to identify multiple texts if they thought that the information in the sentence came from multiple sources. On average, the raters identified the source of information as coming from 2.1 documents, with a range of 0 to 8. The percentage of agreement between the raters was calculated by using a criterion that, for each sentence, if any of the documents chosen by one of the raters agreed with any of the documents chosen by the second rater, it was considered that the two raters agreed on the source. Using this method, the agreement between the raters was 63%. The fact that the agreement between the raters is not that high is not surprising for this type of task. Many of the documents contain very similar pieces of information, since all are on the same topic but often just differ on their interpretation of the same historical events. In addition, because of the total length of the texts (6,097 words), it required a great deal of effort on the part of the raters to locate the correct information.

Since the raters picked an average of two documents for each sentence, the best two matches by LSA for each sentence were used for making predictions. The percentage of agreement between the raters' predictions and LSA's predictions was calculated in the same manner as it had been between the two raters. The agreement between each rater and LSA was 56% and 49%. While not as high as the interrater agreement, the fact that the LSA predictions can get within 7% of the agreement between raters indicates that LSA is able to make many of the same predictions as those of the raters. Considering that the task required LSA to pick two documents out of the set of 21, the method is still performing well above the chance level of picking the documents randomly, which would be 9.5%.

The approach of using LSA to predict the source of a subject's knowledge on the basis of what a subject wrote shows promise. By analyzing the sentences from a subject's summary of information from a set of texts, the method can predict which documents are reflected in the sentences. These predictions are close to those made by human raters. The fact that the interrater agreement between the raters and also between the raters and LSA is fairly low indicates the difficulty of the task due to the high degree of semantic similarity of information across the documents.

Characterizing the Quality of Essays

The above results indicate that LSA can perform matching on the basis of semantic content. For characterizing the quality of essays, one can think of the degree of semantic similarity between what was read in the texts and what was written in the essay as a measure of how much information was learned from the texts. Thus, subjects who write higher quality essays should have captured more of the semantic information from the texts.

Unlike the first experiment, which just used the information on *which* text was the most similar, this experiment used information on *how* semantically similar what a subject wrote in an essay was to what the subject had read. Recall that the result of using an LSA analysis on the essays is that it returns a rank-ordered list of the match-

ing sentences in the original texts on the basis of the cosine between the vectors of the two texts. For grading essays, Foltz et al. (1994) used this cosine measure as a characterization of the quality of the essay. The more similar sentences in an essay are to the original texts, the higher the score. This approach serves as a measure of retention of information. It reflects the degree to which, in their essays, subjects can recall and use the information from the texts.

The same 24 essays used in the previous experiment were used here. Four graduate students in history, who had all served as instructors, were recruited. After becoming familiarized with the 21 texts that the subjects had read, they graded the essays on the basis of what information was cited and the quality of the information cited, using a 100-point scale and also assigning a letter grade from A through F. They were instructed to treat the grading of the essays much in the same way as they would for the undergraduate classes they had taught. In addition, the graders had to read through the original 21 texts and choose the 10 most important sentences in the texts that would be helpful in writing an essay.

Two measures of the quality of the essays were computed using LSA. The first examined the amount of semantic overlap of the essays with the original texts. Each sentence in each essay was compared with all sentences in the original texts, and a score was assigned on the basis of the cosine between the essay sentence and the closest sentence in the original texts. Thus, if a subject wrote a sentence that was exactly the same as a sentence in the original text, they would receive a cosine of 1.0, while a sentence that had no semantic overlap with anything in the original texts would receive a cosine of 0.0. A grade was assigned to the subject's essay on the basis of the means of the cosines for all the sentences in the essay. While this measure captures the degree to which the semantic information in the subject's essay is similar to that of the original texts, the measure could be considered a measure of plagiarism or rote recall. If a subject wrote sentences that were exactly the same as the original texts, the assigned grade would be very high.

The second measure determined the semantic similarity between what a subject wrote and the 10 sentences the expert grader thought were most important. In this way, it captured the degree of overlap with an expert's model of what was important in the original texts. For this analysis, a grade was assigned to each essay on the basis of the mean of the cosines between each sentence in the essay and the closest of the 10 sentences chosen by the expert grader.

The grades assigned for the essays by the graders were correlated between the graders and also with the two measures of the quality of the essays. The correlations are shown in Table 1. The correlations between graders ranges from .367 to .768, indicating that there was some variability in the consistency of the grades assigned by the graders. Most particularly, Grader 4, who had had the least experience in grading essays, did not correlate as well with the other 3. The grades assigned by the first LSA

Table 1
Correlation of Grades Between Expert Graders
and the Two LSA Prediction Methods

Grader	Grader				LSA	
	1	2	3	4	Overlap With Texts	Expert Model Match
1		.575*	.768*	.381	.418†	.589*
2			.582*	.412*	.552*	.626*
3				.367	.317	.384‡
4					.117	.240

* $p < .01$. † $p < .05$. ‡ $p < .06$.

measure (overlap with texts), correlated significantly with 2 of the 4 graders. Thus, grades assigned by human graders do depend greatly on whether the essay captures a lot of the semantic information of the original texts. The grades assigned by the second LSA measure (overlap with expert model) correlated well with those of 3 of the 4 graders, and the correlations were stronger than the first measure. So, the quality of an essay can be characterized as a match between the expert's model of the domain and what was written in the essay. Indeed, the graders' correlations with the LSA expert model are well within the range of the correlations between the graders.

The results indicate that LSA is a successful approach to characterizing the quality of essays and grading done by LSA is about as reliable as that of the graders. Calculating the amount of similarity between what was read and what was written provides an effective measure of the amount of learning by the subject. The results also have implications for understanding what is involved in grading essays. The LSA expert model results indicated that up to about 40% of the variance in subjects' essays can be explained by just the amount of semantic overlap of sentences in the essays with the 10 sentences in the texts that a grader thinks are important. Thus, graders may be looking to see if the essay cites just a few important pieces of information. Additional investigations are currently being performed, both at an application level to refine the LSA approach as a method for grading essay exams and at a theoretical level to determine what degree of subjects' grades are based on the semantics of what was written in contrast to other such factors as the quality of the writing and the ability to write a coherent essay.

Measurements of Text Coherence

The LSA method can also be used for a very different type of analysis used in text comprehension, the measurement of coherence. Propositional overlap measures of textual coherence have been found to be an effective method of predicting the comprehensibility of text (Kintsch & Vipond, 1979). In such a measure, the coherence of the text is calculated by examining the repetition of referents used in propositions through the text (e.g., van Dijk & Kintsch, 1983). This calculation of propositional overlap can be performed at both a local level and a global level for the text. The degree of repetition of the arguments is highly predictive of the reader's recall (Kintsch & Keenan, 1973). For example, readers with low knowl-

edge of a domain will succeed best with a text that is highly coherent (McNamara, Kintsch, Songer, & Kintsch, 1996). Thus, a propositional analysis of the text can suggest places in the text where the coherence breaks down and will affect the reader's recall. Repairs to these places can then improve overall comprehension (Britton & Gulgoz, 1991).

Like propositional models, LSA can measure the amount of semantic overlap between adjoining sections of text to calculate coherence. Foltz et al. (1993) applied LSA to make coherence predictions on a set of texts developed by McNamara et al. (1996). They revised a text on heart disease into four different texts by orthogonally varying both the amount of local coherence (by replacing pronouns with noun phrases and adding descriptive elaboration) and macro coherence (by adding topic headers and paragraph link sentences). One of the four texts was then read by subjects, after which their comprehension was assessed.

To apply LSA to making predictions of coherence on the four texts, LSA first was trained on the semantics of the domain, the heart. A semantic space was derived by performing an SVD analysis on 830 sentences from 21 encyclopedia articles about the heart. This resulted in a 100-dimensional semantic space of 2,781 unique words. Then, taking each text, the LSA predictions on the coherence of the text was made by calculating the amount of semantic overlap between adjoining sentences in the text. Thus, for each text, the cosine distance was found between the vector of sentence *N* to the vector for sentence *N*+1. The mean of all the cosines for a text was then calculated to generate a single number representing the mean coherence for a text. The resulting mean cosines were 0.177 for the text with low local and low macro coherence, 0.205 for the text with low local and high macro coherence, 0.210 for the text with high local and low macro coherence, and 0.242 for the text with high local and high macro coherence. Thus, the predicted coherence by LSA was incrementally higher for the texts that had more coherence.

A question raised by using LSA for calculating coherence is whether LSA actually captures the semantics of the domain, or whether the number of words overlapping between sentences is just as appropriate a measure? Thus, the LSA predictions were compared with predictions of coherence made by measuring simple word overlap. Word overlap can be calculated in the same manner as LSA, except by using the full number of dimensions of the word by document matrix, rather than by the reduced number generated by the SVD. This provides an equivalent cosine value based on the number of overlapping words between sentences. The resulting coherence predictions based on word overlap were 0.155 for the text with low local and low macro coherence, 0.150 for the text with low local and high macro coherence, 0.152 for the text with low macro and high local coherence, and 0.162 for the text with high local and high macro coherence. These results indicate that, compared with the LSA predictions, there is very little difference between the

four texts in terms of word overlap. This suggests that when the texts were revised to improve coherence, the revisions were more at a semantic level than just a repetition of words across sentences. Thus, LSA captures coherence effects on the basis more generally of semantic similarity rather than just shared words among sentences.

The results of the LSA and word overlap predictions were also compared with results generated by McNamara et al. (1996) on subjects' comprehension after reading the text. Figure 1 shows the mean predicted coherence versus the mean of the subjects' comprehension for the four texts. A regression between the four data points for the predicted coherence and the subjects' comprehension was significant for the LSA measure ($r^2 = 0.853$, $p < .05$) but not for the word overlap measure ($r^2 = 0.098$). Thus, the predictions made by LSA are consistent with propositional models which predict better comprehension for texts with higher levels of coherence. The fact that word overlap did not correlate with comprehension indicates that comprehension cannot be predicted by a word-overlap model of coherence.

The coherence predictions made by LSA can also be applied to other applications that use texts, such as document segmentation. In document segmentation, the goal is to determine boundaries within a text where the topic shifts. These boundaries can then identify separate segments of texts that cover different topics. Analyses from Foltz et al. (1993) have indicated that areas in texts where coherence is very low tend to be places where the topic shifts. By identifying breaks in coherence, the method can be used to divide a text into discrete sections. Document segmentation has many applications for presenting on-

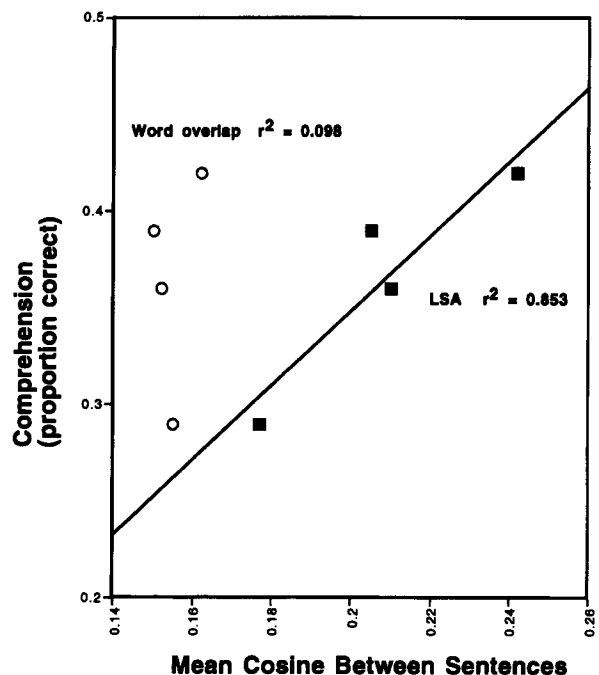


Figure 1. Comprehension versus coherence for LSA and word overlap measures.

line information by breaking up large texts into more manageable units (Hearst, 1995).

DISCUSSION

LSA has a variety of applications to text-based research. The ability to match at a semantic level pieces of textual material allows LSA to perform analyses that were formerly done only through hand-coding. Results comparing LSA's predictions with hand-coding indicate that the percent agreement between LSA and humans is close to the percent agreement between human coders.

The comparisons made by LSA are similar to approaches that use propositions in that they both make comparisons at a semantic level rather than a surface feature level. However, the grain size is slightly larger for LSA than for propositions. Propositions typically represent semantic information at a clause level, while LSA is more successful in performing analyses at a sentence or paragraph level. The few words in a clause make the vectors in LSA highly dependent on the words used in that clause, whereas sentences contain enough words to permit a vector that more accurately captures the semantics of the sentence.

LSA does require a large amount of text in order to perform the SVD analysis. Greater amounts of text help define the space by providing more contexts in which words can co-occur with other words. Typically, 200 contexts (e.g., sentences or paragraphs) would be the minimum needed. Nevertheless, if a text to be analyzed is fairly short, the space can be augmented by adding additional texts on the same topic. Our research has shown that on-line encyclopedia articles are a good source of these texts. LSA also requires a significant amount of processing power, and most analyses are currently performed on UNIX workstations. A typical SVD scaling takes a few minutes on a workstation, and the LSA comparisons between pieces of text take just a few seconds. By comparison, the equivalent hand-coding of propositions can take several hours or days, depending on the size of the text. Given that the processing speed and memory capacities of desktop personal computers have improved greatly, LSA analyses should be able to be performed on desktop machines as well.

In conclusion, research in text comprehension often requires experimenters to perform a variety of analyses on textual material. LSA appears to be a promising application for these researchers. The method is automatic and fast, permitting quick measurements of the semantic similarity between pieces of textual information.

REFERENCES

- BOVAIR, S., & KIERAS, D. E. (1984). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 315-362). Hillsdale, NJ: Erlbaum.
- BRITT, M. A., ROUET, J. F., GEORGI, M. A., & PERFETTI, C. A. (1994). Learning from history texts: From causal analysis to argument models. In G. Leinhardt, I. L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 47-84). Hillsdale, NJ: Erlbaum.
- BRITTON, B. K., & GULGOZ, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83, 329-345.
- DEERWESTER, S., DUMAIS, S. T., FURNAS, G. W., LANDAUER, T. K., & HARSHMAN, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, 391-407.
- DUMAIS, S. T. (1990). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments, & Computers*, 23, 229-236.
- FOLTZ, P. W., BRITT, M. A., & PERFETTI, C. A. (1995, January). *Measuring text influence on learning history*. Paper presented at the 5th Annual Winter Text Conference, Jackson, WY.
- FOLTZ, P. W., KINTSCH, W., & LANDAUER, T. K. (1993, July). *An analysis of textual coherence using latent semantic indexing*. Paper presented at the Third Annual Conference of the Society for Text and Discourse, Boulder, CO.
- GRAESSER, A. C. (1981). *Prose comprehension beyond the word*. New York: Springer-Verlag.
- HEARST, M. A. (1995). TileBars: Visualization of term distribution information in full text information access. In *Proceedings of the Conference on Human-Computer Interaction, CHI '95* (pp. 59-66). New York: Association for Computing Machinery.
- KINTSCH, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 95, 363-394.
- KINTSCH, W., & KEENAN, J. (1973). Reading rate and retention as a function of the number of the propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274.
- KINTSCH, W., & VIPOND, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L.-G. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-365). Hillsdale, NJ: Erlbaum.
- MCNAMARA, D. S., KINTSCH, E., SONGER, N. B., & KINTSCH, W. (1996). Text coherence, background knowledge and levels of understanding in learning from text. *Cognition & Instruction*, 14, 1-44.
- PERFETTI, C. A., BRITT, M. A., ROUET, J. F., GEORGI, M. C., & MASON, R. A. (1994). How students use texts to learn and reason about historical uncertainty. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 257-283). Hillsdale, NJ: Erlbaum.
- RATCLIFF, R., & MCKOON, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning & Verbal Behavior*, 17, 403-418.
- TURNER, A. A., & GREEN, E. (1978). *Construction and use of a propositional textbase*. (From JSAS catalogue of selected documents in psychology, p. 1713)
- VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

(Manuscript received November 13, 1995;
revision accepted for publication November 28, 1995.)