

Reasoning from multiple texts: An automatic analysis of readers' situation models

Peter W. Foltz
Department of Psychology,
Box 30001, Dept. 3452
New Mexico State University
Las Cruces, NM, 88003
pfoltz@crl.nmsu.edu.

M. Anne Britt
Department of Psychology
226 Vincent Science Hall
Slippery Rock University
Slippery Rock, Pa 16057
meb@sruvm.sru.edu

Charles A. Perfetti
Learning Research and Development
Center
University of Pittsburgh
Pittsburgh, PA, 15260
perfetti@vms.cis.pitt.edu

Abstract

In reading multiple texts, a reader must integrate information from the texts with his or her background knowledge. The resulting situation model represents a rich elaborated structure of events, actions, objects, and people involved in the text organized in a manner consistent with the reader's knowledge. In order to evaluate a reader's situation model, a reader's summary must be analyzed in relation to texts the subject has read as well as to more general knowledge such as an expert's knowledge. However, this analysis can be both time-consuming and difficult. In this paper, we use an automatic approach called Latent Semantic Analysis (LSA) for evaluating the situation model of readers of multiple documents. LSA is a statistical model of word usage that generates a high-dimensional semantic space that models the semantics of the text. This paper describes three experiments. The first two describe methods for analyzing a subject's essay to determine from what text a subject learned the information and for grading the quality of information cited in the essay. The third experiment analyzes the knowledge structures of novice and expert readers and compares them to the knowledge structures generated by the model. The experiments illustrate a general approach to modeling and evaluating readers' situation models.

Introduction

In order to comprehend a text, a reader must integrate the information contained in the text with his or her background knowledge of the world. This integration, or situation model (e.g., van Dijk & Kintsch, 1984), is a rich elaborated structure of events, actions, objects, and people involved in the text organized in a manner consistent with the reader's knowledge of the domain. In the domain of history, a reader will typically read multiple accounts of the same historical event in order to generate an understanding of the event. These texts can include primary sources, participant's and historian's accounts, and textbooks. The task for the reader is to then integrate this information into a coherent cognitive representation.

Studying reasoning from history texts provides a realistic approach to examining learning from texts. In the real world, texts we read will be related to other texts we have

read, as well as to experiences and knowledge we have acquired earlier. Nevertheless, the research approach raises a host of discourse processing issues that are not typically found in studies of learning from individual texts. These issues include: Do readers form an integrated situation model of the texts or are texts represented separately? Do certain texts have more influence on the reader's situation model than others? What features of texts are the major sources of influence? How do experts and novice situation models differ? Although we can not answer these questions decisively in this paper, we demonstrate some approaches to these questions through evaluating readers' situation models.

To evaluate a reader's situation model, it is necessary to derive a representation of the reader's knowledge. Typically, a reader provides written summaries, takes tests assessing their knowledge, or is asked to judge relationships between concepts. The primary theoretical approach has been to develop cognitive models of the reader's representation of the text (e.g., van Dijk & Kintsch, 1983; Kintsch, 1988). In such a model, semantic information from both the text and the reader's summary can be represented as propositions. This permits the experimenter to make a comparison of the semantic content contained in the text to 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. Nevertheless, propositionalizing texts can be very time consuming and require a lot of effort, often limiting the size of texts that are analyzed.

In this paper, we describe an automatic method that analyzes texts and generates a semantic space that captures many of the semantic associations found in a reader's situation model. The method, Latent Semantic Analysis (LSA) can be applied in the field of text comprehension to evaluate a reader's situation model, providing results similar to propositional analyses. This paper describes some approaches to analyzing the essays of readers of multiple texts as well as demonstrating that the representation of LSA is similar to that generated by readers of the texts.

Latent Semantic Analysis (LSA) is a statistical model of word usage that models semantic relationships between pieces of textual information. A brief technical overview of LSA will be provided here, while more complete

descriptions of LSA may be found in Deerwester Dumais, Furnas, Landauer & Harshman (1990) and Foltz (In press). 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 could be considered also to 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.

In order to analyze a text, LSA first generates a matrix of occurrences of each word in each document (sentences or paragraphs). LSA then decomposes the matrix using singular-value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis. The SVD 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. An advantage of the approach is that matching can be performed between two pieces of textual information, even if they have no words in common. To illustrate, if LSA were trained on a large number of documents, including:

- 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, LSA automatically captures a 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 sets of textual information, we can determine the semantic similarity between them.

In recent years, a variety of research approaches to generating high-dimensional semantic spaces have developed models that capture semantic meaning based on analyzing large amounts of textual information (e.g., Landauer & Dumais, 1994; Lund, Burgess & Atchley, 1995; Schutze, 1992). LSA has been used in a variety of applications for

representing semantic knowledgebases for modeling results from text and memory experiments. Landauer and Dumais applied LSA for modeling the semantic associations necessary for taking vocabulary tests and for predicting results from studies of semantic priming (See also Lund, Burgess & Atchley for a related approach). Foltz, Kintsch & Landauer (1993) and Foltz (In press) modeled the coherence of texts with LSA to make predictions of reader's comprehension. This paper describes results from three experiments that use LSA to evaluate readers' situation models. The first two experiments analyze subjects' essays to determine from what text a subject learned the information and to grade how much relevant information is cited in the essay. The third experiment analyzes the knowledge structures of subjects and compares them to the knowledge structures generated by LSA.

Predicting the source of the subjects' knowledge

In research on the subject's reasoning after reading multiple documents, it is 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 subjects' 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 6097 words. After reading the texts, 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 most influence in the subjects' essays. In this experiment, we reanalyzed the essays, using LSA to predict 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 likely indicate the source of the subject's knowledge.

To perform the LSA analysis, the texts were first run through the SVD scaling to generate a semantic space on the topic of the Panama canal. The 21 texts the subjects read (6097 words), along with paragraphs from 8 encyclopedia articles on the Panama Canal (~4800 words) and excerpts from 2 books (~17000 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 Panama Canal related words to help better define the semantics of the domain. The LSA analysis resulted in a 100 dimensional space made up of 607 text units by 4829 unique words.

For the analysis of the essays, the vector for each sentence from each subject's essay was compared against 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 best matched along with a cosine indicating the degree of similarity on a scale of 0 to 1. For example, performing 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 returned would be the following two sentences:

MF.2.1 Nov. 2, 5:30 P.M.: U.S.S. Nashville arrives in Colon Harbor with 42 marines. (cosine: 0.640)

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: 0.556)

The codes at the beginning of 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 semantic similarity. 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.

In order to determine the effectiveness of LSA's predictions about the source of each subject's knowledge, they were compared against the 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 could identify multiple texts as the sources for any sentence. On average, the raters identified the source of information as coming from 2.1 documents, with a range of 0 to 8. The percent 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, then it was considered that the two raters agreed on the source. Using this method, the agreement between the raters was 63 percent. The fact that the agreement between two humans was not that high is not surprising. Many of the documents contained similar pieces of information, since all were on the same topic but often just differed on their interpretation of the same historical events.

Since the raters picked on average two documents for each sentence, the best two matches by LSA for each sentence were used for making predictions. The percent agreement between the raters' predictions and LSA's predictions was calculated in the same manner as between the two raters. The agreement between each rater and LSA was 56 percent and 49 percent. While not as high as the inter-rater agreement, the fact that the LSA predictions can get within 7 percent of the agreement between raters indicates that LSA is able to make many of the same predictions as those of the raters.

By analyzing the sentences from subjects' summary of knowledge gained from a set of texts, LSA can predict which

documents are reflected in their sentences. These predictions are close to those made by human raters. This permits a characterization of which texts are having the greatest influence on a subject's situation model as reflected in their summary.

Characterizing the quality of essays

Grading can be characterized as a process of determining whether a reader's situation model is appropriate compared to a text that was read or compared to the situation model of a grader. 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 is what a subject wrote in an essay to what the subject read. For grading essays, the cosines between a subject's sentences and sentences in the original texts were used as a characterization of quality of the essay. The more similar sentences in an essay are to the original texts, the higher the score. Thus, this approach can be thought of as a measure of retention of information. It should reflect the degree to which subjects can recall and use the semantic information from the texts they read in their essay.

The same 24 essays as in the previous experiment were used. Four graduate students in history who had all served as instructors and teaching assistants were recruited. After becoming familiarized with the 21 texts that the subjects had read, they graded the essays based on what information was cited and the quality of the information cited, using a 100 point scale. They were instructed to treat the grading of the essays much in the same way as they would for undergraduate classes they had taught. In addition, the graders read through the original 21 texts and choose the ten most important sentences that were in the texts that would be helpful in writing the 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 against all sentences in the original texts and a score was assigned based on 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 based on 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, it can be thought of as 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 ten sentences the expert grader thought were most important. In this way, it captures the degree of overlap with an expert's model of what is important in the original texts. For this analysis, a grade was assigned to each essay based on the mean of the cosines between each sentence in the essay and the closest of the ten 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.

	Grader 1	Grader 2	Grader 3	Grader 4
Grader 2	.575 **			
Grader 3	.768 **	.582 **		
Grader 4	.381	.412 *	.367	
LSA overlap with texts	.418 *	.552 **	.317	.117
LSA expert model match	.589 **	.626 **	.384 †	.240

**p<.01 * p<.05 † p<.06

Table 1. Correlation of grades between expert graders and the LSA methods

The correlations between graders ranges from .367 to .768, indicating some variability in the consistency of the grades assigned by the graders. The grades assigned by the first LSA measure, which examined the amount of overlap with texts, correlated significantly with two of the four graders. This indicates that the grades assigned by human graders 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, which captures the similarity between what a subject wrote and the sentences that the expert grader thought were most important, correlated well with three of the four graders and, overall the correlations were stronger than the first LSA measure. This indicates that 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 LSA expert model are well within the range of the correlations between the graders, demonstrating that the automatic grading done by LSA was about as reliable as that of the graders. Additional analyses using a weighted mean grade of the experts can improve LSA's correlation of grades to experts to 0.68, which is as good, if not better than the graders correlation with each other.

The results indicate that the representation generated by LSA is sufficiently similar to the readers' situation model to be able to characterize the quality of their essays. Calculating the amount of similarity between what was read and what was written provides a good 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 percent of the variance in subjects' essays can be explained by just the amount semantic overlap of sentences in the essays with 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.

While LSA characterizes the degree to which the subjects' situation model matches that of the text of an expert, there still remain questions as to the degree to which other factors such as the quality of the writing and the ability to write a coherent essay are correlated with or are independent of the subjects' situation model.

Comparing reader's knowledge structures to LSA's semantic distances

One of the assumptions of using LSA to model subjects' situation models is that the semantic structures generated by LSA's analysis of the text correspond to the knowledge structures of the readers of the text. The successful results from the first two experiments indicate that these semantic structures do capture useful features of the reader's representation. The third experiment provides a more direct investigation of these knowledge structures by having subjects make explicit ratings of the semantic similarity between concepts mentioned in the texts and comparing them to the semantic similarity predicted by LSA.

Nineteen undergraduates read the same texts on the Panama canal as in the above experiments. After reading the texts, they were presented with a list of 16 concepts mentioned in the texts. The concepts covered a wide range of issues from the text, including people (President Roosevelt, U.S. Marines), events (U.S. recognizes Panama), and key objects and concepts from the story (The right of transit, The Bidlack Treaty). The subjects rated all 120 possible pairs of the concepts as to how related the two concepts were on a 7 point scale. In addition, the two experts from Experiment 1 both performed the rating task. Each expert performed the task twice, separated by a one month interval, in order to characterize how stable their knowledge of the domain was. Predictions of the similarities between concepts were also made by LSA by determining the cosine between each pair of concepts.

To determine the similarity in knowledge structures, the similarity ratings between concepts were correlated between the experts, the novice subjects and LSA. The two experts had correlations with themselves of 0.86 and 0.63, indicating some variability in even the experts' characterization of the relationship between concepts. (For all analyses, any correlation above 0.18 is significant at the .01 level.) The correlations of ratings between the experts ranged from 0.39 to 0.62, while the novices correlated with themselves with an average of 0.26 and a range from -0.12 to 0.48.

Because of the variability in both the experts' and the novices' ratings, two sets of ratings were derived, one representing a general expert's knowledge structures, by averaging the experts' ratings, the other representing a general novices' knowledge structures by averaging the novices' ratings. The correlation of the average expert rating with the average novice rating was 0.75, while the correlation of LSA with the average expert rating was 0.41 and with the average novice rating was 0.36. The fact that the correlation of LSA to the humans is not as high as the correlation between the experts and novices is not surprising. Averaging the human judgments removes a lot of the variability in their data, but since the LSA predictions

were based on a single set of judgments, no averaging could be done. Thus, there would likely be more variability in the LSA predictions, resulting in lower correlations with humans than the correlations between the averaged humans. While the LSA correlation is not as strong as the correlation between the experts and novices, LSA manages to capture many of the same semantic relationships as those identified by both the experts and the novices. The above correlations also indicate that LSA's representation is somewhat closer to that of an expert in the domain than that of a novice.

In addition to calculating correlations between the LSA and the expert and novice ratings, Pathfinder analyses (Schvaneveldt, 1990) were performed on the similarity matrices for the experts, novices and LSA. The Pathfinder analyses derives a network structure which represents concepts as nodes and distances between concepts as the number of intervening links between the nodes. Network similarity scores, which indicate the degree to which two networks share similar links on a scale of 0 to 1, were computed. The comparison of pairs of networks indicated that the expert and novice networks were more similar than expected by chance ($SIM=0.48$, $ExpectedSIM=0.14$, $p<.01$) as well as the expert and LSA networks ($SIM=0.32$, $ExpectedSIM=0.14$, $p<.01$). However, the LSA network was not significantly similar to the novice network ($SIM=0.21$, $ExpectedSIM=0.14$, $p>.1$). As in the correlation results, these results indicate that the structures produced by LSA were similar to those produced by the experts.

Since the correlation and network similarity scores between LSA and the experts were lower than those between experts and novices, an analysis of the pairs of concepts was done in order to determine on which pairs LSA was making predictions that differed greatly from the experts. The experts' averaged ratings and the LSA cosines were converted to Z-scores and the differences between Z-scores was computed. This permitted a characterization of overpredictions, where the model predicted that relationship between two concepts was closer than the actual predictions made by the experts, and underpredictions, where the model predicted that relationship between two concepts was not as close as the predictions made by the experts.

The Z-score analyses indicated that LSA tended to overpredict the relationship between pairs of concepts when the two concepts occurred exclusively together in the texts or in cases when the two concepts shared terms (e.g., "U.S. Marines" and "U.S. recognizes Panama"). This indicates that LSA's representation of the texts was sometimes overly contextually based. Given that the LSA scaling was based on only a small sample of text (607 text units), there would be certain words that would occur in only one context and thus would not have as rich a semantic representation. This problem could be alleviated by providing a larger training set for LSA, such as using more encyclopedia articles, which would provide more contexts for words to occur.

The underpredictions made by LSA occurred in cases when the concepts represented more global events and players such as "President Roosevelt" and "U.S. recognizes Panama" as well as "President Roosevelt" and "U.S. Marines". These pairs were never described together in the text, yet their

relationships can be inferred if a reader has enough general knowledge about the role of presidents in fomenting revolutions or having command over the Marines. Since LSA was only trained on semantic associations about Panama, it lacked much of this deeper situational knowledge. Thus, in this case, LSA had a somewhat limited situation model in that it was effective at representing semantic relationships about the Panama canal situation, but did not have more domain general history knowledge as do both the experts and novices. Again, this problem could be remedied by providing a larger training set for LSA, such as using more encyclopedia articles covering a wide range of topics in history.

Discussion

When reading multiple texts a reader must integrate the information across the texts as well as combine it with their previous knowledge. LSA captures this integration of information, representing concepts in a semantic space in which vector similarity between concepts represents a characterization of semantic relatedness. The primary properties of LSA that permit this representation is through performing an analysis of co-occurrence of words and then a statistical reduction of the analysis to capture high level associations between terms and documents. Thus, LSA captures a feature inherent in textual information: words that tend to co-occur together or tend to occur in the same contexts will be more semantically related. Based on the results of these studies, the statistical approximation of semantic relatedness generated by LSA corresponds to the knowledge structures generated by readers of multiple texts. In this manner, LSA has similar properties to that of a reader's situation model. It generates a semantic representation of textual information that integrates the information across texts. In addition, it can also integrate that information with outside information (e.g., previous knowledge) as long as the model has been trained on that information. Based on this representation, LSA can be used as a tool for evaluating readers' summaries and as a way of modeling readers' knowledge structures.

The first two experiments demonstrate that LSA provides a representation in which comparisons can be made between readers' summaries of their knowledge and the original texts they have read. These comparisons permit a variety of characterizations of a reader's situation model. By comparing sentences in a reader's summary to sentences from the texts that were read, LSA provides a measure as to which texts had the greatest influence on the reader. In addition, the degree of semantic similarity between the reader's summary and the original texts provides a measure of the amount of information the reader learned from the texts. By comparing the reader's summary to sentences that an expert thought was important, LSA provides a measure of the degree to which the reader's situation model matches that of the expert. Since essays provide a rich source of information about a reader's knowledge, but are often difficult to analyze, this approach permits automatic analyses of the reader's knowledge.

Because the derivation of the semantic space is based on an analyses of multiple texts, LSA generates a rich semantic

representation. In the case of the above experiments, the space was derived based on about 23000 words, although the same method has been applied to corpora with millions of words (e.g., Landauer & Dumais, 1994). The third experiment demonstrates that distances between concepts in the semantic space correspond to conceptual distances in the readers' representation of the texts. This indicates that LSA captures a deep associational structure between concepts that is similar to the reader's situation model of the texts. The types of errors that LSA made in its predictions of concept similarities show that both novices and experts are incorporating information in the text with their global knowledge of political roles. Since LSA was not trained with texts that provide this information, it lacked a global a situational representation compared to humans. Thus LSA is dependent on the type of texts provided in order to successfully mimic a human situation model. For the present experiments, the situation model generated was limited only to knowledge about the Panama Canal situation. With a larger corpus, such as that used in the Landauer and Dumais research, LSA would have a more general representation, capturing more global associations that are made by humans when integrating information from the text with their background knowledge. Future research will address these issues as, well as continue to evaluate the boundaries of where the representation generated by LSA successfully captures or differs from the reader's representation.

In summary, LSA can serve both as a tool and as a modeling technique for text researchers. As a tool it can be applied as a method to analyze readers' summaries and characterize both the source and quality of a reader's knowledge. As a model of readers' comprehension, it can be used to generate a semantic representation that captures important features of a reader's situation model.

Acknowledgments

This work has benefited from contributions from Tom Landauer, Walter Kintsch, Susan Dumais, and Mara Georgi. Support for this research was partially funded by a grant to the Learning Research and Development Center from the Office of Educational Research and Improvement, Department of Education.

References

- 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: Lawrence Erlbaum Associates.
- 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.
- Foltz, P. W. (In Press). Latent Semantic Analysis for text-based research. *Behavior Research Methods, Instruments & Computers*.
- 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.
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 95, 363-394.
- Landauer, T. K. & Dumais, S. T. (1994). Memory model reads encyclopedia, passes vocabulary test. Talk presented at the Psychonomics Society.
- Lund, K, Burgess, C. & Atchley, R. A. (1995) Semantic and Associative Priming In High-Dimensional Semantic Space. In *Proceedings of Cognitive Science*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schutze, H. (1992). Dimensions of Meaning. *Proceedings of Supercomputing*.
- Schvaneveldt, R. W, (1990) *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex Publishing.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.