Behavior Research Methods, Instruments, & Cognition
1999, 31 (2), 208-214

THEORIES AND APPLICATIONS OF HIGH-DIMENSIONAL SEMANTIC MODELS

Automatically deriving readers' knowledge structures from texts

PETER W. FOLTZ and AMBER D. WELLS
New Mexico State University, Las Cruces, New Mexico

Latent semantic analysis (LSA) serves as both a theory and a method for representing the meaning of words based on a statistical analysis of their contextual usage (Foltz, 1996; Landauer & Dumais, 1997). In experiments in the domains of psychology and history, we compared the representation of readers' knowledge structures of information learned from texts with the representation generated by LSA. Results indicated that LSA's representation is similar to readers' representations. In addition, the degree to which the reader's representation is similar to LSA's representation is indicative of the amount of knowledge the reader has acquired and of the reader's reading ability. This approach has implications both as a model of learning from text and as a practical tool for performing knowledge assessment.

The acquisition of knowledge requires both the acquisition of a set of concepts related to that knowledge and an understanding of the relationships among those concepts. Together, the concepts and the relationships among them form a representation of the learner's knowledge structure of a topic. By assessing a learner's characterization of the relationships among concepts, we can measure that person's knowledge structures. In addition, correlations of these relationships among multiple learners can be used to characterize the similarity of the knowledge structures between participants. This approach permits a researcher to analyze the effect a particular text may have on a reader's knowledge structures, which may in turn be used to determine what characteristics of the text had particular effects on learning. In this paper, we present a method of automatically deriving a knowledge structure based on statistical analysis of a text and assess how well this knowledge structure corresponds to the knowledge structures formed by a human reader of that text.

The development of semantic models of memory has long relied on using psychometric approaches and a large number of participants' judgments to determine the relationships among concepts (see, e.g., Osgood, Suci, & Tannenbaum, 1957). With the advent of more powerful computing, as well as the availability of online corpora, new techniques have been developed for automatically deriving high-dimensional semantic representations (Foltz, 1998). These techniques have permitted the development of semantic models without having to collect large numbers of human judgments and can be used to provide cognitively plausible representations of human knowledge structures. Two such models, hyperspace analog to language (HAL; Burgess, Livesay, & Lund, 1998) and latent semantic analysis (LSA; Landauer & Dumais, 1997), have been widely used to model semantic knowledge representations. In the research described below, we use LSA for modeling readers' knowledge structures based on the texts they have read.

LSA serves as both a theory and a method for representing the meaning of words based on a statistical analysis of their contextual usage (Foltz, 1996; Landauer & Dumais, 1997). The full details of LSA will not be provided here, but may be found in Landauer and Dumais (1997) and Landauer, Foltz, and Laham (1998). On the basis of an analysis of large amounts of textual information, LSA generates a high-dimensional semantic space in which words are represented as vectors in the space. The similarity of two words can be compared in the semantic space through determining the cosine or dot product between the vectors. This representation permits a comparison of semantic similarity among words, as well as larger units of text, even if those words are used in different contexts. The representation of the semantic space captures effects due to the pattern of word usage across many contexts by abstracting out key semantic similarities between words, based on the correlation of word usage across contexts (e.g., sentences or paragraphs). For example, consider the following two sentences:

The authors thank Anne Britt, Mara Georgi, and Charles Perfetti for assistance with the first experiment, and Sharon Bailey for her assistance in developing materials and running the second experiment. They also thank Frank Durso, Adrienne Lee, and two anonymous reviewers for helpful comments. Correspondence should be addressed to P. W. Foltz, Department of Psychology, Department 3452, Box 30001, New Mexico State University, Las Cruces, NM 88003 (e-mail: pfoltz@nmsu.edu).

> To retrieve a specific memory, you first need to identify one of the strands that leads to it, a process called *priming*.

and

> William James referred to *priming* as the "wakening of associations."

In an LSA analysis, the word *priming* will tend to have a similar vector as words such as *associations*, *James*, and *memory*, since it occurred in the same context as those words. In addition, the vector for *associations* will also be similar to the vector for *memory*, since they both occur in the context of *priming*, although they never occur together in the same context (sentence). Thus, LSA captures not just the associations of words within contexts but also higher order relationships among words based on the words with which they tend to co-occur.

LSA has been tested using a variety of different approaches to demonstrate that its representation of meaning corresponds well to the representations generated by humans. These tests have included representing and evaluating students' knowledge of history as expressed in their essays (Foltz, 1996; Foltz, Britt, & Perfetti, 1996), predicting the results of lexical priming and performance on word-sorting tasks (Landauer & Dumais, 1997), predicting results of word-sorting tasks (Laham, 1997), and modeling the effects of text coherence on comprehension (Foltz, Kintsch, & Landauer, 1998). In each case, LSA was first used to derive a semantic space by training LSA on a corpus of text, thereby providing LSA with extensive examples of words across a wide range of contexts. This extensive training is necessary to provide LSA with sufficient language exposure so that it can model relationships among words. The texts used for the training were either texts relevant to the domain that was going to be tested (e.g., psychology textbooks) or large amounts of general text (e.g., an encyclopedia). The representation generated by LSA was then compared with that of humans on the same task.

One approach to examining similarities in knowledge representations has been through the use of proximity data. Proximity data, in which a participant rates the relatedness of concepts, provide measures of the psychological similarity of those concepts and can be used to determine a participant's conceptual model of relationships among a group of concepts. There have been a variety of approaches for modeling with proximity data, including pathfinder networks (Schvaneveldt, 1990; Schvaneveldt, Durso, & Dearholt, 1989), multidimensional scaling (Shepard, 1974), and hierarchical clustering (Johnson, 1967). The research on pathfinder networks has more explicitly examined the representations of both novices and experts on the basis of proximity ratings (e.g., Rowe, Cooke, Hall, & Halgren, 1996). In such studies, proximity ratings of both domain novices and experts are collected in order to evaluate how the novices' knowledge structures become more similar to the experts' with experience. In a similar vein, we can evaluate LSA's proximity ratings of concepts in semantic spaces and compare them with the proximity ratings of participants, we can measure the degree to which LSA's representation corresponds to the participants' representation.

In this paper, we further evaluate LSA by comparing readers' and LSA's characterization of semantic relatedness among concepts. This approach permits a determination of how well LSA's representation of the similarity between vectors for concepts, based on the analysis of a text, corresponds to the representation generated by a reader of the same text. Additionally, LSA's representation can be compared with the domain representation of human readers, both before and after exposure to a particular text, to determine whether the human knowledge structures change to more closely resemble LSA after reading the text. This provides a measure of the influence of a text on a reader's knowledge structures. In addition, by assessing the reader's knowledge through other types of tests, we can determine how well LSA's assessment of a reader's knowledge structures corresponds to more traditional tests of a reader's knowledge. Finally, the approach also assesses whether LSA represents only information due to just co-occurrence of the concepts within contexts, or whether the higher order associations derived by LSA account for the representation.

Two experiments were performed in which the overall procedure was to have students read a text, in either history or psychology, and then rate the similarity of pairs of concepts from the text. The similarity ratings were used as representations of the readers' knowledge structures. In both experiments, domain experts and novices performed the rating task. The results from the domain experts provide some measure of the knowledge structures of participants with prior knowledge of the topic, while the results from the novices provide a characterization of how the text influenced their knowledge representation. By comparing the two with LSA, we could determine whether LSA's representation is more similar to the representation of participants with more or less prior knowledge. After exposure to a text, human readers will have generated some knowledge representation of the information presented in the text. Readers who are experts with respect to the domain would be expected to develop a more complete as well as more complex knowledge representation of the material. Domain experts should have the requisite background knowledge to understand the interrelationships among the ideas presented, while domain novices would not be expected to be as sensitive to these interrelationships. Our claim here is that after exposure to the text, LSA develops a complex representation of the textual information because LSA is sensitive to higher order relationships among concepts and not just local co-occurrence of terms. Thus, one would expect that the representation generated by LSA would look more like that of experts than that of novices.

EXPERIMENT 1

Method

Participants

The participants were 21 students at the University of Pittsburgh enrolled in an introductory psychology course. All students participated for course credit. In addition, two professional researchers who were experts in the domain participated on a voluntary basis.

Materials

Text. Twenty-one excerpts from texts related to the United States involvement in the Panamanian revolution were collected. These texts were used originally in Perfetti, Britt, Rouet, Georgi, and Mason (1994). The total length of the texts was 6,097 words in 424 sentences. The text was presented to the participants as a printed booklet.

Rating task. Sixteen concepts were chosen that were central to the events of the Panamanian revolution and that were discussed in the texts. These concepts included "U.S. Marines," "President Roosevelt," "Colonel Shaler," and "USS Nashville." A booklet was created that presented pairs of the concepts. Participants were instructed to rate the relatedness of each pair of concepts on a scale that ranged from 1 to 7. Each booklet contained all 120 possible pairings of concepts printed in a randomized order.

Procedure

Participants first read the booklet of texts on the Panamanian revolution. The participants were given 45 min to read the booklet. They were encouraged to spend the whole time reading and to go back and reread the booklet if they had additional time. After reading the booklet, participants rated the similarity of 120 word pairs consisting of all possible pairs of the 16 terms. Participants were instructed to make their ratings on the basis of the "overall relatedness" of the word pairs. The ratings were made on a 7-point scale (1 = *unrelated*; and 7 = *highly related*). Finally, all participants took the reading comprehension portion of the Nelson–Denny test and completed a short general history test in which they had to circle names of historical figures from a list of names. The two domain experts also performed the same rating task on the 120 pairs of concepts without having read the text. The experts performed the rating task twice, with a week separating their trials, in order to obtain a measure of internal reliability of the ratings.

Results

LSA Analysis

In order to derive a representation of knowledge related to the Panamanian revolution, a semantic space was developed on the basis of an analysis of the original texts (424 sentences) read by the participants. In addition, 47 paragraphs from encyclopedia articles about the history of the Panama Canal and 136 paragraphs from two books on the history of the Panama canal were included in the development of the space. These extra paragraphs provided additional examples of the contexts of words and concepts that were in the texts read by the participants. LSA could have just been trained on the original 424 sentences seen by the participants, but the original material gave very few occurrences of each term. The power of LSA's analysis relies on it being able to generalize how a term occurs in multiple contexts. It might be argued that the participants in the study have had extensive exposure to language and thus are experts with respect to most concepts in the text. The reading of the text just helps define novel relationships among the concepts. The additional training material provided to LSA merely helped provide more information about the meaning of the concepts provided in the text by providing some additional exposure to the concepts.

The LSA analysis resulted in a matrix of 4,829 unique terms by 607 contexts and was represented using 100 dimensions.[1] The occurrence of terms in the matrix was weighted by the log*entropy of their frequency.[2] On the basis of this semantic space, concepts could be compared with each other by determining the cosine between the vectors for each of these concepts in 100 dimensions. Concepts that consisted of multiple terms were represented as a single vector that was the centroid of the vectors for each term. The cosine of the angle between the vectors was then used as the measure of relatedness for the two concepts.

Analysis of Correlations Among Ratings

For analyzing the similarity of readers' knowledge structures, the 120 concept pairs rated by each participant were correlated with the concept pairs rated by other participants or rated by LSA. For 120 pairs, an r of greater than .18 was significant at the .05 level and an r of greater than .24 was significant at the .01 level. A Fisher's r to z was performed on all correlations before correlations were averaged.

The correlations between the expert raters provide some measure of agreement on the knowledge representation of the domain. Since the experts performed the rating twice, there were four pairings of ratings between the experts in which the correlations ranged from .39 to .62, with an average of .51 ($p < .01$). In addition, the correlations for each expert from the first rating task to the second were .86 for 1 expert, and .63 for the other. These results provide an upper bound on expected correlations between LSA and the human raters.

A comparison of the performance of the novices on the rating task indicated that the average correlation between novices was .26 ($p < .01$), with a range of $-.12$ to .48. This indicates some general agreement on the knowledge structures. The experts' ratings also significantly correlated with the novices' ratings, with an average correlation of .37 for 1 expert and .29 for the other expert. This result indicates that the novices were deriving some of the same knowledge structures as those of the experts.

Through comparing the novice and expert ratings with the similarity measures derived by LSA, we can determine how well the representation generated by LSA corresponds to that of the experts and novices. The average correlation of LSA's ratings to the novices' ratings was .19 ($p <$ .05), with a range of $-.15$ to .4. LSA's average correlation to 1 expert was .31 ($p < .01$) and to the other expert, the correlation was .37 ($p < .01$).

A second way of examining the correlations among LSA, the experts, and the novices is to examine the mean of the ratings for the novices and correlate that with LSA's and the experts' ratings. Note that taking the mean of the ratings and then correlating them removes much of the noise introduced by individuals and therefore provides a more stable overall characterization of the novices' representation as a whole. The means of the novices' ratings were correlated with those of each of the individual experts' ratings, and then the mean of this correlation was taken. This mean correlation between experts and novices was $r = .59$ ($p < .01$).[3] LSA correlated with the mean of the experts with $r = .41$ ($p < .01$) and with the novices with $r = .36$ ($p < .01$). A comparison of correlated correlation coefficients (see, e.g., Meng, Rosenthal, & Rubin, 1992), however, showed no significant difference between the correlations for the experts and the novices, on the one hand, and LSA ($Z = 0.66$, $p = .25$), on the other.

Finally, we can examine the novices' performance on the Nelson–Denny reading comprehension test and the history knowledge test in order to determine whether the novices' correlation to LSA is related to their abilities in reading or to their general knowledge of history. Each novice's ratings of the concepts were correlated with the ratings of LSA. This correlation for each novice was then used as a score to indicate how similar the novice's ratings were to the ratings of LSA. Assuming that LSA provides an accurate model of the representation of the text, novices who correlate more strongly with LSA should also perform better on the external measures of their history knowledge and reading performance. Thus, these scores were then correlated to the novices' performance on the two tests. The novices' degree of similarity to LSA's score correlated significantly to their performance on the history test ($r = .31$) as well as to their performance on the reading comprehension test ($r = .37$). These results indicate that students with higher abilities in history and higher reading ability derived a representation from the text that was more similar to LSA's representation than those participants who had lower knowledge or reading abilities. Thus, LSA's representation captures some effects of the quality of the novices' knowledge.

In order to determine where LSA's representation differed greatly from that of the experts, a Z-score analysis was performed. The LSA concept pair cosines and the experts' averaged ratings were transformed to Z scores and their differences were computed. This permitted a determination of where LSA greatly overpredicted or underpredicted the relatedness of concepts. Using a casual classification, the results indicated that LSA tended to underpredict relatedness on concepts that required a high-level understanding of relationships in history. For example, the largest underprediction by LSA was of the relationship between "President Roosevelt" and "U.S. Marines." These two concepts are treated very separately in the text, with the marines intervening in Panama, while Roosevelt dealt with treaty negotiations. Nevertheless, humans readers inferred the relationship between the president and the actions of the marines on the basis of their general world knowledge, even though this relationship was not made explicit in the text. Had LSA been trained on a much larger corpus—for example, an encyclopedia—it might have inferred the relationship between the marines and the president, but in the limited corpus, these two concepts tended to occur in very different contexts. LSA's overpredictions tended to occur with concepts that shared words, such as "rights of sovereignty" and "rights of transit." *Rights* is a fairly low-frequency term, so the vectors for concepts containing "rights" tended to be more similar than would be predicted by humans.

Discussion

Experiment 1 illustrated that LSA captures some of the effects of relatedness among concepts expressed in the text. The representation generated by LSA based on an analysis of the texts corresponded both to the representations of the novices and to those of the experts. In addition, LSA's representation corresponded better to the representations of the novices who were either more skilled in history or had better reading comprehension skills. These results indicate that LSA is capturing an effective representation of the text and that this representation reflects the participants' abilities in extracting the appropriate knowledge from the texts. Experiment 2 was designed to assess whether LSA can be used to determine changes in knowledge structures as students learn from a text, as well as to further investigate how the knowledge structures generated by LSA correspond to other measures of student ability. Also, a different domain was chosen in order to assess LSA's effectiveness in learning a representation of psychology concepts.

EXPERIMENT 2

Method

Participants

The participants were 40 undergraduates enrolled in an introductory psychology course at New Mexico State University. In addition, 6 graduate students in the psychology department who had worked as teaching assistants for introductory psychology participated as domain experts.

Materials

Text. Participants read a chapter on memory from Myers's (1995) introductory textbook. The chapter was photocopied, single sided, and stapled together to make a booklet. The chapter was 16,863 words in length.

Test questions. Fifteen multiple-choice questions and five short-answer questions were created on the basis of the information from the chapter on memory. The multiple-choice questions were all derived from the test bank associated with the textbook. The questions specifically tested knowledge about the 16 terms taken from the text.

Rating program. Participants completed a rating task on 16 terms taken from the chapter on memory. The task was presented on a Macintosh computer using HyperCard. The program first presented the participants with the 16 terms, and participants were asked to rate their degree of familiarity with the terms. Next the

Table 1
Correlation of Readers' Representation With Latent Semantic Analysis Representation Before and After Exposure to the Text

	Graduate Students	Undergraduates
Before reading text	.31†	.19*
After reading text	.37†	.27*

*$p < .05$; †$p < .01$.

program presented participants with all possible pairs of the 16 terms one at a time in random order. Participants were asked to rate how related the two words in the pair were on a 7-point scale.

Procedure

The undergraduate participants were divided into two groups. One group of 21 participants first completed the rating task at the computer. Participants rated their degree of familiarity with the 16 terms taken from the textbook chapter. Ratings were made on a 7-point scale (1 = *unfamiliar* and 7 = *very familiar*). After completing the familiarity ratings, participants rated the similarity of 120 word pairs consisting of all possible pairs of the 16 terms. Participants were instructed to make their ratings on the basis of the "overall relatedness" of the word pairs. The ratings were made on a 7-point scale (1 = *unrelated*; 7 = *highly related*). After they completed the ratings, the participants were given a photocopy of the textbook chapter to read. They were given 50 min to read the chapter and were reminded of the time after 15 min, after 30 min, and at the 45-min point. Participants were encouraged to use the full 50 min and to become as familiar as possible with the text material during the allotted time. At the end of 50 min, the textbook chapter was removed and participants were directed to the computer, where they again completed the rating task. At the conclusion of this second set of ratings, participants answered 15 multiple-choice and 5 short-answer questions about the text material. Participants were allowed as much time as they needed to answer the questions.

For the second group of 19 undergraduate participants, the participants first read the textbook chapter for 50 min, then completed the rating task. Finally, participants completed the reading comprehension portion of the Nelson–Denny reading test. Participants were given a maximum of 20 min to complete the test. The 6 graduate student participants first performed the rating task, then read the chapter for 50 min, and then performed the rating task again.

Results

LSA Analysis

A semantic space on introductory psychology was developed through performing an LSA analysis of the entire Myers (1995) textbook. This textbook was the same textbook that the participants were using in their course. The semantic space was based on 4,903 paragraphs by 19,160 unique terms and was represented in 300 dimensions. Log*entropy weighting was used for the terms in the LSA analysis. As in Experiment 1, the relatedness between concepts was measured through determining the cosine between the vectors representing the concepts in the semantic space.

Analysis of Correlations Among Ratings

As in Experiment 1, the ratings by the undergraduates, graduate students, and LSA were correlated with each other in order to determine similarities in representations. The ratings given by participants after they had read the text were examined; the average correlation among the undergraduates was .17 ($p > .05$), and the average correlation among graduate students was .45 ($p < .01$). Thus, as in Experiment 1, there was much greater agreement in the knowledge representation of those with more expertise in the domain.

To compare the representation generated by LSA, LSA's ratings were correlated to the mean undergraduate and graduate ratings. In addition, the mean of the undergraduate ratings was correlated with the individual ratings of each graduate student, and then the mean of that correlation was taken. The correlation of LSA to undergraduates was $r = .27$ ($p < .01$), and to graduate students it was $r = .37$ ($p < .01$). However, a comparison of correlated correlations indicated that there was not a significant difference between the undergraduate and graduate correlations with LSA ($Z = 1.13, p = .13$). The correlation between graduate and undergraduates was $r = .47$ ($p < .01$).

In order to examine the role of learning from the text, we compared the ratings performed by participants who had rated concepts both before and after the reading of the text. The mean ratings of both undergraduates and graduate students were correlated with the ratings generated by LSA. Table 1 shows the average correlation of the undergraduate and graduate students with LSA before and after reading of the text. A test of correlated coefficients, however, showed that for both the undergraduates and graduate students, there was not a significant change in the correlations from before to after reading the text (undergraduates, $Z = 1.17, p = .12$; graduates, $Z = 1.27, p = .10$). This lack of significant change may be due to the fact that most participants were already familiar with the topic, and the text did not have a strong enough effect to greatly change their knowledge structures.

As in Experiment 1, the degree to which each undergraduate's representation correlated with LSA was compared with the tests of the participants' reading abilities or topic knowledge. The tests of the 21 participants who took the multiple-choice and short-answer test on memory topics were graded, and a grade from 0 to 20 was derived. The overall scores for their correlation with LSA were significantly correlated with their performance on the memory test ($r = .58, p < .01$). For the 19 participants who took the Nelson–Denny reading comprehension test, scores for their correlation with LSA were significantly correlated with their performance on the comprehension test ($r = .50, p < .01$).

These results illustrate that the representation generated by LSA corresponds to the representation of the readers of the text and that LSA corresponds better to readers with greater reading skills and knowledge. Nevertheless, it might be argued that the power in LSA's representation is not due to the inductive properties of LSA in capturing higher order associations among concepts, but just to the fact that certain terms tend to co-occur within the same contexts (e.g., paragraphs), and the ratings just reflect whether these concepts tended to occur together. Thus, an additional set of ratings was generated using the full

Table 2
Correlation of Readers' Representation With a Within-Paragraph Similarity Measure Before and After Exposure to the Text Using the Full-Dimensional Space

	Graduate Students	Undergraduates
Before reading text	.08 ($p = .42$)	−.04 ($p = .71$)
After reading text	.15 ($p = .10$)	.02 ($p = .79$)

number of dimensions to represent the semantic space, rather than the reduced set of 300 dimensions. In this case, this is equivalent to setting the number of dimensions to the number of contexts from the original text. This approach has the effect of accounting for term co-occurrence within contexts, while removing LSA's inductive properties of associating terms that occur in similar contexts. Table 2 shows the correlations of the ratings generated by the full semantic space compared with the graduate and undergraduate ratings made before and after reading of the text. All four correlations from Table 1 were significantly greater than those in Table 2. These results illustrate that without the reduced dimensional space, LSA does not account well for the human rating data. Thus, the power behind the representation generated by LSA can be attributed greatly to the inductive properties captured in co-occurrence of terms across contexts.

GENERAL DISCUSSION

The results from these two experiments indicate that LSA provides a representation of the semantic relatedness among concepts that are acquired by readers of a text. Some evidence suggests that the representation generated by LSA is more similar to the representation of domain experts than to that of novices. Although there was not a significant difference in correlations of LSA to experts and novices, the results did show that with greater amounts of domain knowledge and reading skills, the readers' representations became more similar to LSA's representation. Thus better readers and more knowledgeable readers are able to extract the appropriate information in order to build a better knowledge representation. Experiment 2 did not show a significant change in the readers' knowledge structures from before to after reading of the text. This may have been due to the fact that many of the participants were already familiar with the text. Finally, similarity of the representation generated by LSA to the human readers' representation is due to more than just within-paragraph term co-occurrence. LSA is modeling higher order effects of word similarity, even when the words do not appear in the same contexts.

LSA's representation nevertheless does not match the readers' representation perfectly. Novices and experts still tended to correlate more strongly with each other than they correlated with LSA. One primary factor to explain this effect is that, because LSA has been trained only on the text read by the participants, but does not have a history of all previous knowledge of the participants, it can only partially model the effects that any particular text will have on a reader. Thus, without a model of the reader's prior knowledge structures, it assumes that knowledge gained is represented only in terms of the knowledge expressed in the text, but not prior knowledge. A possible solution to this problem would be to use much larger training spaces, which would provide a more general representation of a participant's prior knowledge, and then add in the target text and note how it causes changes in the representation. For example, one could train LSA on a representative set of texts that a participant would have encountered in his/her lifetime and then add in the one target text (see Landauer et al., 1998, for research on such large corpora). Given that the representation generated by LSA is based on only a single exposure to the same text as that seen by participants, LSA does a good job of modeling the relationships among concepts for that text. However, further work must be done in order to determine in what systematic ways LSA's representation deviates from the human representation and how to incorporate the effects of prior knowledge into LSA's representation.

LSA as a Practical Tool

Since LSA can model readers' knowledge representations, it has practical applications for knowledge assessment. For example, the method can be used for performing automatic evaluations of students' knowledge of particular topics within a training system. Through an LSA analysis of a text or set of texts, LSA's representation of relatedness among concepts should correspond well to that of a domain expert. Comparisons can then be made between a student's performance on those concepts and the model. Using this approach, we can generate measures that determine the overall comprehension of information in a domain, as well as methods for diagnosing knowledge deficits on specific concepts within a domain. This information can then be used to determine what additional specific information must be provided to a student in order to improve his/her knowledge of the domain.

For text and educational researchers, LSA can be used as a tool for evaluating how particular texts will influence knowledge structures. Through an analysis of a text, a researcher could predict the relationships among concepts and thus determine whether the representation of information expressed in the book corresponds to a desired representation. This would permit people to select texts on the basis of how well a book corresponds to their preferred representation of the knowledge.

LSA as a Model

The results from these experiments illustrate that LSA can successfully model the acquisition of knowledge from texts. As a reader reads a text, new concepts are acquired and are represented in long-term memory. Associations among these newly acquired concepts are not only a result of the co-occurrence of concepts in similar contexts (e.g., paragraphs), but are also due to higher level associations among the concepts occurring across the contexts. These results are consistent with previous findings showing that LSA can account for many factors involved in the acquisition and representation of meaning (e.g., Landauer & Dumais, 1997; Landauer et al., 1998).

Overall, LSA can model the effects of acquiring new concepts and knowledge structures. This approach permits one to model the effects that different texts on the same topic will have on readers' knowledge structures. In addition, it can make explicit predictions about how certain concepts will be understood in relation to other concepts on the basis of an analysis of the texts. LSA thus serves as an effective model of the effects of texts on learning.

REFERENCES

BURGESS, C., LIVESAY, K., & LUND, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, **25**, 211-257.

FOLTZ, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, **28**, 197-202.

FOLTZ, P. W. (1998). Computational approaches to semantic knowledge representations. *Discourse Processes*, **25**, 127-130.

FOLTZ, P. W., BRITT, M. A., & PERFETTI, C. A. (1996). Reasoning from multiple texts: An automatic analysis of readers' situation models. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Cognitive Science Conference* (pp. 110-115). Hillsdale, NJ: Erlbaum.

FOLTZ, P. W., KINTSCH, W., & LANDAUER, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, **25**, 285-307.

JOHNSON, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241-254.

LAHAM, D. (1997). Latent semantic analysis approaches to categorization. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Cognitive Science Conference* (p. 979). Hillsdale, NJ: Erlbaum.

LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

LANDAUER, T. K, FOLTZ, P. W., & LAHAM, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, **25**, 259-284.

MENG, X. L., ROSENTHAL, R., & RUBIN, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, **111**, 172-175.

MYERS, D. G. (1995). *Psychology* (4th ed.). New York: Worth.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

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.

ROWE, A. L., COOKE, N. J., HALL, E. P., & HALGREN, T. L. (1996). Toward an on-line knowledge assessment methodology: Building on the relationship between knowledge and doing. *Journal of Experimental Psychology: Applied*, **2**, 31-47.

SCHVANEVELDT, R. W. (1990). *Pathfinder associative networks studies in knowledge organization*. Norwood, NJ: Ablex.

SCHVANEVELDT, R. W., DURSO, F. T., & DEARHOLT, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249-284). New York: Academic Press.

SHEPARD, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, **39**, 373-421.

NOTES

1. The number of dimensions chosen must be less than the minimum of the number of contexts and the number of terms in order to get effects beyond co-occurrence. Typically, 100–300 dimensions provides fairly effective representation of the information (see Landauer & Dumais, 1997).

2. The log*entropy weighting function is an information-theoretic approach that tends to downweight terms that occur frequently, both within contexts and across contexts. It has been widely used in the evaluation of information retrieval techniques as well as other psychological tests of LSA.

3. A different approach would be to correlate the mean of the novices ratings to the mean of the expert ratings. However, to make the analysis comparable to the ratings given by LSA, the mean of the correlations between each expert and the mean of the novice ratings are used. In this manner, the similarity ratings made by LSA can be thought of as an expert rating as well. Thus we are comparing how well the individual experts correlate to the mean of the novices in the same way as we are comparing how well LSA correlates with the mean of the novices.

(Manuscript received October 15, 1998;
revision accepted for publication February 19, 1999.)