How Well Can Passage Meaning be Derived without Using Word Order? A Comparison of Latent Semantic Analysis and Humans

Thomas K. Landauer, Darrell Laham, Bob Rehder, and **M. E. Schreiner**
Department of Psychology & Institute of Cognitive Science
University of Colorado, Boulder
Boulder, CO 80309-0345
{landauer, dlaham, rehder, missy}@psych.colorado.edu

Abstract

How much of the meaning of a naturally occurring English passage is derivable from its combination of words without considering their order? An exploratory approach to this question was provided by asking humans to judge the quality and quantity of knowledge conveyed by short student essays on scientific topics and comparing the inter-rater reliability and predictive accuracy of their estimates with the performance of a corpus-based statistical model that takes no account of word order within an essay. There was surprisingly little difference between the human judges and the model.

In the studies reported here, experts were asked to read short student essays about scientific topics with the goal of determining how much knowledge was accurately reflected in a given essay. We measured the readers' success by how well their ratings agreed with each other and how well they predicted scores on an objective test on the same subject.

All current accounts of human discourse understanding assume significant reliance on syntactic structure within sentences and order-dependent relations between sentences. Therefore, if the order of words in the input were randomly scrambled the ability to judge how much correct knowledge it expresses presumably would be significantly reduced.

On the other hand, even the best current methods in automatic information retrieval (IR) use little or no syntactic information in representing documents, relying primarily on "bag-of-words" methods (Harman, 1994). One can conclude from this either that the success of IR proves that "bags-of-words" are ordinarily sufficient to characterize the content of discourse or that the far from perfect performance of such methods proves much is lost by ignoring word order.

Of course, it is obvious that processes that depend on word order and syntax often play important roles in the comprehension of sentences, and that sentence order and inter-sentence coherence relations often have important effects on discourse comprehension. However, it is difficult to know just how much people's extraction of information in ordinary discourse for ordinary purposes depends on such processes over and above what is derived from the combination of lexical items alone. For example, in discourse about a focused semantic domain, speakers or writers may create few sentences that could not, with sufficient time and effort, be correctly construed in context even if their internal word order were scrambled. And, given enough effort, a human or machine might be able to properly rearrange or otherwise process disordered sentences.

One way to gain insight into this issue is to compare the performance of a computational method that does not use word order with that of humans when both are posed with the same comprehension-demanding problems. If the computational method can do as well as humans, its input must be sufficient. However, if it fails to equal humans, it could be only because its analysis or representations are defective. Thus, the exploration requires the availability of a computational method that does a reasonable job of mimicking human comprehension. As outlined next, the recently developed Latent Semantic Analysis technique meets this requirement. In addition, such an exploration requires that solutions to the problems posed to human and machine depend on a significant level of discourse comprehension and that the measure of success be such that any source of better or more complete comprehension is reflected. We believe that estimates of the quantity and quality of knowledge conveyed, and their correlations with other measures of knowledge satisfy this requirement adequately for the purpose. Even if such estimates fail to reflect some products of comprehension, for example aesthetic or emotional qualities, they pose a sufficient challenge to comprehension to provide an illuminating test of the sufficiency of the input.

Latent Semantic Analysis (LSA) is a corpus-based statistical method for inducing and representing aspects of the meaning of words and passages reflected in their usage (Berry, Dumais & O'Brien, 1995; Landauer and Dumais, 1996, 1997).[1] It is related to but different from some other corpus-statistic methods (cf. Lund & Burgess, 1995, in press; Schütze, 1992). In LSA a representative sample of text is converted to a matrix of word-types by passages. Cell entries are the frequency of a given word in a given passage. After a preliminary information-theoretic weighting of cell entries, the matrix is submitted to singular value decomposition (SVD) (see Berry, 1992) and a 100-1500

[1] LSA can also be construed as a theoretical model of human acquisition and representation of knowledge (Landauer and Dumais, 1997) Its learning mechanism is equivalent to a particular kind of linear neural network.

dimensional abstract "semantic space" is constructed in which each original word and each original (and any new) passages are represented as vectors. LSA's representation of a passage is just the average of the vectors of the words it contains independent of their order.

It is essential to note that the similarities derived by LSA are not simple co-occurrence statistics. The dimension-reduction step constitutes a form of induction that can extract a great deal of added information from mutual constraints among a large number of words occurring in a large number of contexts. This is important because received wisdom in the language sciences often holds that co-occurrence cannot explain meaning (e.g. Chomsky, 1965). However, traditional arguments against statistical learning of meaning have usually assumed the direct use of surface co-occurrence relations rather than their use as input data for a method of global constraint satisfaction such as LSA.

LSA has proven able to closely mimic several properties of human verbal meaning. Its first success was in improving "bag-of-words" IR by allowing queries to correctly match documents of similar meaning with which they shared no words and to reject documents of the wrong meaning that did contain some query words (see Deerwester, Dumais, Furnas, Landauer & Harshman, 1990; Dumais, 1991, 1994). More recent applications have addressed LSA's ability to represent word and passage meaning more directly. For example, Landauer and Dumais (1996, 1997) found that after training on a student encyclopedia (or, more recently, a corpus of newspaper text), LSA chose the same answers on a standardized English synonym test as did successful foreign candidates for U. S. Colleges. LSA's vocabulary growth per paragraph of text was similar to that of grade-school children, and its learning depended strongly on induction; LSA with optimal dimension reduction was at least three times as effective as ordinary co-occurrence measures.

LSA-based measures have also been found to reflect the relations between individual words and overall passage meaning as evinced in semantic priming experiments (Landauer and Dumais, 1997), and accurately mirrored the sentence-to-sentence coherence of passages and their resulting comprehensibility (Foltz, Kintsch and Landauer, 1993, in press). In addition, LSA measures of the similarity between student essays and instructional text have been found to predict how much the student will learn from the text (Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch & Landauer in press; Rehder, Shreiner, Wolfe, Laham, Landauer & Kintsch, in press). These results show that LSA captures significant portions of the meaning not only of individual words but also of whole passages such as sentences, paragraphs and short essays.

We will first expand somewhat on the description of LSA, then give details of the experiments and results. Finally, some ancillary results that tend to substantiate and clarify the results are presented.

Latent Semantic Analysis.

LSA relies on singular value decomposition (SVD) of the matrix of words-by-contexts derived from a corpus of natural text that expresses human knowledge in a domain of interest. The advantage of SVD is that the linear factorization of which it consists can impose strong constraints on fitting data to model, that the degree of constraint—the number of dimensions used—can be conveniently varied, and that it is computationally feasible for the large datasets sometimes needed to emulate the knowledge sources relied upon by human learners. SVD also produces a natural measure of the similarity between any two entities in its solution space as the cosine of the angle between their vectors, and of the intensity of a single entity as the length of its vector.

Example of text data: Titles of Technical Memos

c1: *Human* machine *interface* for ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

X =	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Figure 1: A word by passage matrix, ***X***, formed from the titles of five articles about human-computer interaction and four about graph theory. Cell entries are the number of times that a word (rows) appeared in a title (columns).

Figures 1 and 2 show a miniature example that gives the flavor of the analysis and demonstrates what the technique accomplishes. This example uses as text passages the titles of nine technical memoranda, five about human computer interaction, and four about mathematical graph theory, rather disjoint topics. The original matrix has nine columns, and we have given it 12 rows, each corresponding to a content word used in at least two of the titles. The titles, with the extracted terms italicized, and the corresponding word-by-title matrix is shown in Figure 1.

The linear decomposition of this matrix by SVD is defined as $\mathbf{X} = \mathbf{W}\,\mathbf{S}\,\mathbf{C}'$, where $\mathbf{X}$ is the original matrix, $\mathbf{W}$ and $\mathbf{C}$ are orthonormal matrices with rows standing for words and contexts respectively, and columns containing derived linearly independent dimensions of representation. $\mathbf{S}$

is a diagonal matrix of singular scaling values. If there are enough dimensions, pre-multiplication of the right-hand matrices perfectly reconstructs the original data. However, if some of the dimensions are omitted the reconstruction is a least-squares best approximation. Because this minimization requires the simultaneous accommodation of all the data it constitutes a form of induction.

To illustrate what dimension reduction does to representations of passages, we computed the inter-correlations (Spearman *r*) between each title and all the others, first based on the raw co-occurrence data, then on the corresponding vectors in the two-dimensional reconstruction; see Figure 2. In the two dimensional reconstruction the topical groupings are much clearer.

Correlations between titles in raw data

	c1	*c2*	*c3*	*c4*	*c5*	*m1*	*m2*	*m3*
c2	-0.19							
c3	0.00	0.00						
c4	0.00	0.00	0.47					
c5	-0.33	0.58	0.00	-0.31				
m1	-0.17	-0.30	-0.21	-0.16	-0.17			
m2	-0.26	-0.45	-0.32	-0.24	-0.26	0.67		
m3	-0.33	-0.58	-0.41	-0.31	-0.33	0.52	0.77	
m4	-0.33	-0.19	-0.41	-0.31	-0.33	-0.17	0.26	0.56

0.02
-0.30 **0.44**

Correlations in two-dimensional space

	c1	*c2*	*c3*	*c4*	*c5*	*m1*	*m2*	*m3*
c2	0.91							
c3	1.00	0.91						
c4	1.00	0.88	1.00					
c5	0.85	0.99	0.85	0.81				
m1	-0.85	-0.56	-0.85	-0.88	-0.45			
m2	-0.85	-0.56	-0.85	-0.88	-0.44	1.00		
m3	-0.85	-0.56	-0.85	-0.88	-0.44	1.00	1.00	
m4	-0.81	-0.50	-0.81	-0.84	-0.37	1.00	1.00	1.00

0.92
-0.72 **1.00**

Figure 2. Comparing intercorrelations among vectors representing titles in the original full dimensional source data and in a two-dimensional reconstruction illustrates how LSA changes passage representations.

In all LSA simulations of human meaning relations attempted so far, there has been a significant nonmonotonic function of accuracy on the dimensionality of the space used to represent the word and passage vectors. Our working hypothesis is that a natural dimensionality is determined by some combination of neural processing architecture and statistical properties of the input corpus and that embedding the data observations in spaces with too small dimensionalities causes unnatural distortions, whereas spaces with greater than optimal dimensionalities do not exploit mutual constraints in the data in the way that humans must.

LSA extracts only a single global relation between words and passages, that of similarity as defined by relative position in high-dimensional space. However the vector of a given word is decomposable, so varying components of its meaning may project differentially onto the vectors of the other words in different local contexts. (This is roughly a continuous analog of feature based meaning combination.) Thus, as we will illustrate later, more specific interpretations of similarity relations may emerge in local contexts through recomputation.

Experiments comparing LSA and humans

Experiment 1

In the main experiment, 94 undergraduates at the University of Colorado were asked to write essays of approximately 250 words on the anatomy, function and purpose of the human heart. The essays were given to two professional readers at Educational Testing Service, Inc., who after reading relevant background material and discussing the knowledge that such an essay should contain, independently read each essay and assigned a quality score from 1 to 5 to reflect their estimate of how much the student knew and correctly conveyed about the subject. The students were also given a 40 point short answer test on the same topic (Wolfe, *et al.*, in press).

LSA was first trained on a set of 27 articles relevant to the heart and circulatory system taken from a version of *Grolier's Academic American Encyclopedia*. This produced a 94-dimensional vector for each of 830 sentence-length passages and 3034 unique words. A filtering "stop list" was used to exclude 439 common words from the analysis. Each sentence of an article was considered a text passage for the purpose of creating the semantic space. Next, a vector was computed for each essay by averaging the vectors of all the words contained in it that were represented in the semantic space. LSA was then used to evaluate the essays by two different methods.

Method 1. First the cosine was computed between a target essay vector and each of the other essays. Then the ten most similar essays to the target were identified. Finally, the target essay was given the cosine weighted average of the scores that the humans readers had assigned to the closest ten. This provided the first of two components of the LSA score, a component we interpret as the semantic direction or quality of the essay. The second component for both methods was the vector length of the target essay, which we interpret as the amount of domain relevant information it contains. Table 1 shows the results for the 94 undergraduate essays. A combined score based on both quality and quantity predicted each of the two human readers estimates as well as they predicted each other. The correlation of the LSA assigned scores with the short-answer test scores (external criterion) was somewhat better than the average correlation between the human graders' scores and the short-answer test scores.

Method 2. Here instead of computing the quality of content measure by similarity to essays scored by humans, we computed the cosine between the target essay and a short text on the topic written by an expert, a section on the heart from a college biology textbook. As shown in Table 1, the results were just about as good as in Method 1. The correlation of the external criterion with the LSA assigned scores was again slightly better than the correlation between the external criterion and the human graders assigned scores.

Correlation between	
Method 1	
Two ETS reader scores:	.77
LSA score and ETS reader 1 score:	.68
LSA score and ETS reader 2 score:	.77
LSA score and average ETS score:	.77
Average ETS and external criterion:	.70
LSA score and external criterion:	.81
Method 2	
LSA score and ETS reader 1 score:	.64
LSA score and ETS reader 2 score:	.71
LSA score and average ETS score:	.72
LSA score and external criterion:	.77

Table 1: Heart essay results.

Experiment 2

In another test of the same kind, 273 introductory psychology student essays were analyzed. The students were given ten-minutes to write the essays on one of three topics in psychology—attachment in children, aphasias, and operant conditioning. Essays were read by two people, one the professor or a graduate student teaching assistant, the other one of two advanced undergraduate psychology majors also serving as teaching assistants.

Method 1 was used to evaluate the essays with LSA, which was trained on the textbook used in the course to yield the semantic space. This space was developed using 4904 paragraphs containing 19,153 unique terms. This analysis did not use a stop list of common words. The analysis was repeated using from 2 to 2100 dimensions. The correlation between the LSA Method 1 assigned grade and the average of the 2 graders rose steadily with dimensionality, leveled out around 400-500 dimensions, and began to decline gradually after about 1,500. The results, shown in Table 2 (at 1500 dimensions), are similar to those for the heart essays; the correlation between the LSA assigned score and the average human score was as good as the agreement between the two human readers. The inter-rater reliability for the two graders in this case was not as good as for the ETS graders. Using Method 1, the inter-rater reliability puts a limit on how well LSA can do because LSA bases its score in part on the human scores of those essays nearest to it in the semantic space. The two readers were especially unreliable on their scoring of the 'attachment in children' essays.

These results show that what humans extracted by reading essays in order to judge the knowledge of the authors was not much superior to what LSA extracted for the same purpose. Human understanding neither produced significantly more agreement with another human nor better predicted an external measure of students' knowledge. While this does not prove that the human readers obtained no meaning from the essays that LSA missed, it does show that they derived little if any more total information relevant to one highly meaning dependent task. Given the scientific nature of the material, one might have supposed the extraction of this much information to require consideration of word order and syntax; it apparently did not.

Correlation between	
All Essays (n = 273)	
Two reader scores:	.65
LSA score and average reader score:	.64
Attachment in children (n = 55)	
Two reader scores:	.19
LSA score and average reader score:	.61
Aphasias (n = 109)	
Two reader scores:	.75
LSA score and average reader score:	.60
Operant conditioning (n = 109)	
Two reader scores:	.68
LSA score and average reader score:	.71

Table 2: Psychology essay results.

Ancillary findings

The data from the heart essay experiments were analyzed further. First the words in the student essays were divided into technical and non-technical sets, and the method 2 procedure repeated on each. As shown in Table 3, LSA performance is not based solely on some sort of count of relevant rare words: LSA cosine measures based on either technical or common general vocabulary terms were significantly correlated with human estimates and objective test results. However, as to be expected, because vector length reflects the amount of domain specific knowledge, and common words come from a completely general domain with no special relevance to the question, the only vector length component that was useful was one that included technical words. Combined cosine and vector length scores, based on either technical words only ($r = .72$) or on all words ($r = .77$) were as accurate predicting objective measures as human estimates. Recall from Table 1 that on average the human ratings correlated .70 with the objective test scores.

Second, the quantity measure was examined in more detail. In addition to vector length, we tried simply counting the number of words in the essays, both before and after common words were removed by applying the stop list. Table 3 shows that before stoplisting, there was no correlation between essay word count and either the ETS grades and the external criterion, and after stoplisting the correlations were still quite weak. The vector length however is highly correlated with the human grades and test results. Thus, the preweighting and dimension reduction steps performed by LSA are of crucial importance for representing the human knowledge contained in an essay.

Third, the relations between vectors for all the essays written by students were studied by a supplementary analysis. Similarities (cosines) between every pair of essays were computed and the result subjected to a one-dimension multi-dimensional scaling. The resulting single dimension (see Table 3) has a very clear interpretation as the goodness or sophistication of the essay. Indeed, when position along this dimension is taken as the LSA quality measure, the predictions of both human judgments and the short-answer test scores are almost the same as the measures based on cosines with a standard text. We believe that the reason that the single dimension reflects goodness of answer is simply that the goal of the essay question writers was to pose a problem that would cause essays to differ primarily in how much correct information they contained. The analysis thus illustrates that the components of LSA vectors can carry particular, contextually interpretable meaning.

	Correlation with	
	Average ETS Score	External Criterion
Tech. vs. Nontech. Words		
Cosine essay•standard for tech words:	.59	.65
Cosine essay•standard for nontech. words:	.47	.53
Cosine essay•standard for all words:	.63	.68
Essay Word Count		
Before stoplisting:	.03	-.01
After stoplisting:	.25	.16
Vector length:	.65	.65
Multidimensional Scaling		
One-dimension score:	.59	.66
Cosine with expert text:	.62	.68

Table 3: Further analyses of heart essay measures.

Conclusions

This paper presents new evidence that a great deal of information about the meaning of passages may be carried by words independently of their order. Students were asked to write short essays that would demonstrate their knowledge of scientific topics. The amount and correctness of topic-relevant knowledge displayed in the essays was determined either by judgments of two expert human readers or by measures derived from the text by LSA. The principal findings were (1) that LSA-based measures—which take no account of word order—were as closely related to human judgments as the latter were to each other, and (2) that the LSA measures predicted external measures of the same knowledge as well or better than did the human judgments.

These results and analyses demonstrate that most of the meaning derived by people in reading the texts was also extracted by the LSA learning method without recourse to syntax. It is important to note that LSA's ability to do this is not a simple matter of counting and weighting words or co-occurrences, but depends on its derivation of a semantic space of optimal dimensionality from the mutual constraints reflected in the use of many words in many contexts. The fact that LSA can capture as much of meaning as it does without using word order shows that the mere combination of words in passages constrains overall meaning very strongly.

How can this be? In addition to the contrary theoretical presumptions mentioned earlier, various intuitive and rational arguments suggest that such representations must fall far short of extracting as much meaning from text as do human readers. For instance, the following two sentences are identical for LSA, but have very different meanings for a human reader: "It was not the sales manager who hit the bottle that day, but the office worker with the serious drinking problem."; "That day the office manager, who was drinking, hit the problem sales worker with a bottle, but it was not serious."

Nonetheless, what such examples prove is only that a method that ignores word order cannot always render completely correct comprehension. They tell us almost nothing about the relative contribution of the combination of words in an average utterance and that of their order to what humans usually understand. Indeed, it is entirely possible that humans could find scrambled word order utterances unacceptable or incomprehensible even if the underlying process by which they represent meaning is mostly independent of order. That could happen, for example, because an obligatory input parsing step acts as a gate keeper that rejects ill-formed sentences even when order is actually unimportant for meaning.

A hypothetical language example illustrates the issue in a more general theoretical manner. There would be nothing to prevent one from constructing a language in which all significance is transmitted only by the order-free combination of words in marked-off sets. An infinite number of different messages could be transmitted. Now suppose that for efficiency in processing the language, say for looking up words in a code book, its users decide that message sets should always be ordered alphabetically, and to detect and avoid transmission errors, they decide to reject any message that is not alphabetical. We thus have a language with a very strict and useful syntax in which the syntax has no direct role in the representation of meaning.

We do not mean to imply that the syntax of English or other natural language serves only such purposes. It is obvious, for example, that word order has an important function in representing and transmitting mathematical ideas and those aspects of common knowledge—such as who bit whom—that involve non-commutative logical operators and bracketing. Nonetheless, we have no way of knowing to what extent syntax is a matter of conventional ordering that serves other purposes than meaning representation. A few among many other roles that syntax may play might be: (a) to reduce loads on the human working memory and processes needed to extract meaning from word combinations, or (b) to ease the construction of sequential utterance production from unordered sets of words, or (c) to subserve important matters of aesthetic or social style, or (d) merely to maintain functionless conventions generated by the social equivalence of genetic drift.

Thus, while it is obvious that complex syntactic regularity is well mastered and applied by human language users, we simply do not know how much of the interpretation of differences in meaning between natural utterances needs syntax and how much can be recovered without it. The point is that it was and is premature to conclude that the word constituency of passages cannot be an important, or even usually sufficient, bearer of meaning.

The results reported here show that for at least one task requiring a serious amount of discourse comprehension, word order was not needed for performance equaling that of humans. Apparently, then, the discourse involved did not contain much task-relevant meaning that was in doubt without syntactic interpretation. One reason that this could be the case is that it is extremely difficult to construct complex discourse in which component sentences express correct knowledge, and the sequence of sentences expresses proportionally more, without so constraining the mix of specific and general words that no very different overall meaning could be expressed with the same set of words. But if this is so, then it also follows that a properly constructed device could extract the meaning from the combination of words alone.

Alternatively, it remains a theoretical possibility that humans generate discourse from some abstract ideational representation that is more like a bag or cloud of words than a set of ordered sentences, and that much (but not all) of the order in which words are produced serves other purposes than meaning representation, for example that of choosing a minimum path, maximum information, easily articulated string of words to express the meaning in the original representations, or to produce a string from which receivers with particular processing systems and limitations can conveniently unpack the original. Conceivably, for example, LSA represents meaning in much the same way as humans, but because it is not limited by the same data processing constraints, it does not need word order to reconstruct meaning from input.

In any event, the fact that highly meaning based judgments can be accurately made without using word order both provides promising possibilities for artificial intelligence applications and suggests directions in which important presumptions in theories of human knowledge representation and processing ought to be reconsidered.

References

Berry, M. W. (1992). Large scale singular value computations. *International Journal of Supercomputer Applications*, 6(1), 13-49.

Berry, M. W., Dumais, S. T. and O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM: Review*, 37(4), 573-595.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA.: MIT Press.

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(6), 391-407.

Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments and Computers.*, 23(2), 229-236.

Dumais, S. T. (1994). Latent semantic indexing (LSI) and TREC-2. In D. Harman (Ed.), *National Institute of Standards and Technology Text Retrieval Conference*, NIST special publication.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1993, July). *An analysis of textual coherence using Latent Semantic Indexing*. Paper presented at the meeting of the Society for Text and Discourse, Boulder, CO.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (in press). Analysis of text coherence using Latent Semantic Analysis. *Discourse Processes.*

Harman, D. (Ed.) (1994). *National Institute of Standards and Technology Text Retrieval Conference*, NIST special publication.

Landauer, T. K., & Dumais, S. T. (1996). How come you know so much? From practical problem to theory. In D. Hermann, C. McEvoy, M. Johnson, & P. Hertel (Eds.), *Basic and applied memory: Memory in context.* Mahwah, NJ: Erlbaum, 105-126.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In J. D. Moore & J. F. Lehman (Ed.), *Cognitive Science Society*, Pittsburgh, PA: Lawrence Erlbaum Associates, 660-665.

Rehder, B., Schreiner, M. E., Wolfe, B. W., Laham, D., Landauer, T. K., & Kintsch, W. (in press). Using Latent Semantic Analysis to assess knowledge: Some technical considerations. *Discourse Processes.*

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (in press). Learning from text: Matching readers and text by Latent Semantic Analysis. *Discourse Processes.*