The Embodiment of Amodal Symbolic Knowledge Representations

Max Louwerse, Zhiqiang Cai, Xiangen Hu, Matthew Ventura, Patrick Jeuniaux

Department of Psychology
Institute for Intelligent Systems
University of Memphis
{mlouwers, zcai, xhu, mventura, pjeuniax} @ memphis.edu

Abstract

Latent semantic analysis (LSA) is a statistical, corpus-based technique of representing knowledge. It has been successfully used in a variety of applications including intelligent tutoring systems, essay grading and coherence metrics. From its very introduction LSA has been claimed to simulate aspects of human knowledge representation. This amodal symbolic view immediately resulted in a rebuttal from cognitive scientists who argued that LSA can never come close to human knowledge representation, because it lacks embodiment into perceptual experience or action as in human cognition. The ultimate test to determine the embodiment of LSA is to evaluate how well it operates on typical embodied dimensions like spatiality and temporality. The results of such a test would have an impact on theories of meaning and knowledge representation, and would help us in the development of natural language processing modules for applications that use these embodied dimensions. Six Multidimensional Scaling representations are derived from the LSA cosines between spatial and temporal words. Results show that by combining LSA with MDS previously hidden spatial and temporal relations can now be revealed.

Introduction

Natural-language based knowledge representations (NL-KR) have a number of advantages over knowledge representation systems that are not natural-language based. One of these advantages is that human knowledge is encoded – or can at least be decoded – using natural language. Within these NL-KR systems two approaches can be distinguished. One is application-driven and provides the necessary tools for the computational linguist and is precise in nature. An example of such an application-driven system is WordNet (Miller, 1990), a large semantic database containing information regarding the hypernymy, hyponymy, synonymy and other lexical aspects of words. These lexical features are all carefully handcrafted and make WordNet a reliable and precise resource for a variety of applications, including coherence measurement (Graesser, McNamara, Louwerse & Cai, 2004), question-answering systems (Pasca & Harabagiu, 2001) and logic form representations (Rus & Moldovan, 2002). The other NL-KR system approach is cognition-driven and simulates aspects of human cognition. These systems may be less precise than the application-driven systems, but have some important advantages: They are not hand-crafted, but instead meaning is induced without considerable human assistance; they are flexible and allow for approximate relations between words. An example of such a system is Latent Semantic Analysis.

Copyright © 2005, American Association for Artificial Intelligence (www.aaai.org). All rights reserved

Latent Semantic Analysis

Latent semantic analysis (LSA) is a statistical, corpus based, technique for representing world knowledge that estimates semantic similarities on a scale of -1 to 1 between the latent semantic representation of terms and texts. The input to LSA is a set of corpora segmented into documents like paragraphs or sentences. Mathematical transformations create a large term-document matrix from the input. For example, if there are m terms in n documents, a matrix of $A = \left(f_{ij} \times G(j) \times L(i, j)\right)_{m \times n}$ is obtained. The value of f_{ij} is a function of the integer that represents the number of times term i appears in document j: $L(i; j)$ is a local weighting of term i in document j; and $G(j)$ is the global weighting for term j. Such a weighting function is used to differentially treat terms and documents to reflect knowledge that is beyond the collection of the documents. This matrix of A has, however, lots of redundant information. Singular Value Decomposition (SVD) reduces this noise by decomposing the matrix A into three matrices $A = U\Sigma V'$; where U is m by m and V is n by n square matrices, such that $UU' = I$; $VV'=I$ (an orthonormal matrix), and Σ is m by n diagonal matrix with singular values on the diagonal. By removing dimensions corresponding to small singular values and keeping the dimensions corresponding to larger singular values, the representation of each term is reduced as a smaller vector with only k dimensions. The new representation for the terms (the reduced U matrix) are no longer orthogonal, but the advantage of this is that only the most important dimensions that correspond to larger singular values are kept. Each term now becomes a weighted vector on K dimensions. The semantic relationship between words can be estimated by taking the normalized dot product (cosine) between two vectors (Hu, et al., 2003).

What is so special about LSA is that the semantic relatedness is not (only) determined by the relation between words, but also by the words that accompany a word (Landauer & Dumais, 1997; Kintsch, 1998).

LSA in Knowledge Representation Applications

The method of statistically representing knowledge has proven to be useful in a variety of studies. It has been used as an automated essay grader, comparing student essays with ideal essays (Landauer, Foltz & Laham, 1998) and performs as well as students on TOEFL (Test of English as a foreign language) tests (Landauer & Dumais, 1997).

More recently, LSA has also been used for several other applications. First, it plays an important role in Coh-Metrix (Graesser, McNamara, Louwerse & Cai, 2004; Louwerse, McCarthy, McNamara & Graesser, 2004), a web-based tool that analyzes texts on over 50 types of cohesion relations and over 200 measures of language, text, and readability. LSA measures the semantic relatedness between sentences, paragraphs and texts.

LSA has also been used in intelligent tutoring systems like AutoTutor. AutoTutor engages the student in a conversation on a particular topic like conceptual physics or computer literacy. The system assists students in actively constructing knowledge by holding a conversation in natural language. AutoTutor uses LSA for its model of the world knowledge and determines the semantic association between a student answer, and ideal good and bad answers (Graesser, Wiemer-Hastings et al., 2000). The computational architecture of AutoTutor and its learning gains have been discussed extensively in previous publications (Graesser, VanLehn, et al., 2001).

Recently, intelligent systems have been developed that communicate with a human conversational partner on finding a route on a map, called Intelligent MapTask (iMAP) agents (Louwerse, et al., 2004). This constrained semantic domain triggers a highly interactive, incremental conversation and allows us to carefully measure the interaction of modalities like eye gaze, intonation, dialog structure and gestures. LSA would be an immediate choice for the knowledge representation of the iMAP system but it is unclear how LSA performs on dimensions like time and space. In fact, it may be questionable whether LSA can map space and time in the first place. This issue has also been relevant for the question whether LSA simulates human cognition.

LSA as Knowledge Representation Theory

Since its early days LSA has been considered to provide a solution to Plato's problem, the problem of how observing a relatively small set of events can result in knowledge representations that are adaptive in a large, potentially infinite variety of situations (Landauer & Dumais, 1997). It does this by mapping initially meaningless words into a continuous high dimensional semantic space, more or less simulating cognition (Landauer, 1999). More specifically, a first-order process that associates stimuli (words) and the contexts they occur in (documents). This process is very much like classical conditioning where stimuli are paired based on their contiguity or co-occurrence. These local associations are next transformed by means of SVD into more unified knowledge representations by removing noise. Like language comprehension, memory for the initial local associations (surface structure) becomes memory for more global representations (central meaning). LSA can thereby be seen as a theory of knowledge representation, induction and language acquisition (Landauer & Dumais, 1997; Landauer, 2000; Louwerse & Ventura, 2004).

But many psychologists and cognitive scientists have argued that corpus based models of word meaning can simply not be the whole story. For instance, embodied theorists (Barsalou, 1999; Glenberg and Robertson, 2000) claim that the basis of linguistic and non-linguistic understanding is the sensorimotor experiences of actions. Consequently, associative models using only amodal symbols can never fully identify the meanings of words. It is like learning a language in a foreign country with only a dictionary: Without grounding the words to bodily actions in the environment we can never get past defining a symbol with another symbol.

The response to this embodied view of meaning is simple: LSA simply does not have the advantage of raw perception and motoric intentions and one could predict that if these perceptual symbols were to be included in LSA its performance may be even more impressive than its current record. So a solution to the embodiment problem in LSA is to supply LSA with perceptual elements (Landauer, 1999; 2000). The problem is that "these [perceptually grounded] frames are organized by space and time, not simply by associative strength" and therefore "a simple extension of LSA from words to perception probably will not work" (Barsalou, 1999: 639).

In sum, space and time are the key challenges for LSA, because these are typically embodied and can thus presumably not be modeled in an amodal symbolic knowledge representation. This problem from a cognition-based perspective also has implications for the application-based approach: If LSA cannot deal with space and time it may not be the right NL-KR system for intelligent systems that deal with spatial tasks, like iMAP agents (Louwerse, et al., 2004). In a number of observations we take up this challenge and investigate to what extent LSA can actually simulate spatial and temporal dimensions in an embodied kind of fashion.

Method

As we outlined earlier LSA's knowledge of the world comes from a set of corpora segmented into documents like paragraphs or sentences. All evidence suggests that larger corpora benefit the performance of LSA. For all six studies that follow we therefore selected the (commonly used) Touchstone Applied Science Associates (TASA) corpus that consists of approximately 10 million words of unmarked high-school level English text on Language arts, Health, Home economics, Industrial arts, Science, Social studies, and Business. This corpus is divided into 37,600 documents, averaging 166 words per document.

In all observations we compared the LSA cosine output between a set of selected words by creating a matrix of cosine values. Because we are primarily interested in underlying dimensions of the relationships between these words, these LSA matrices were then supplied to an ALSCAL algorithm to derive a Multidimensional Scaling (MDS) representation of the stimuli (Kruskal & Wish, 1978). That is, we view the matrix of LSA cosine values as a matrix of Euclidean distances. This matrix is compared with arbitrary coordinates in an *n*-dimensional space. The coordinates are iteratively adjusted such that the Kruskall's stress is minimized and the degree of correspondence maximized.

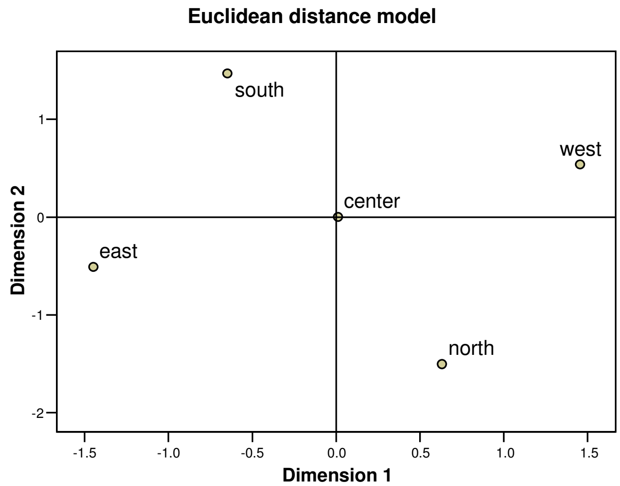
For reasons of clarity, we selected a limited number of items (typically < 10) and plotted these in a two-dimensional space (interpreting five dimensions is awkward).

Because our goal at this point is to detect meaningful underlying dimensions that allow us to explain similarities and dissimilarities between the words and because of the exploratory character of this study, we feel an exploratory statistical measure like MDS is warranted.

Observation 1

In this first observation we investigated the semantic relationships between wind directions North, South, West and East by applying MDS to the LSA similarity matrix of terms. The traditional association between terms will not provide much more information than that these four words have a semantic relation. For instance, *North* and *South* have a strong semantic relation (cos = .74) and so have *West* and *East* (cos = .76). North and West on the other hand have a weaker semantic relation (cos = .11). In sum, we can conclude that there are semantic similarities between these terms. But we would like to take this one step further: If we plot the semantic similarities in a two-dimensional graph, can we obtain spatial information that the *North* and *South* axis is orthogonal to the *West* and *East* axis? Clearly, there is nothing in the co-occurrences and semantic associations that would suggest this orthogonal relation. After computing the cosine values between these terms, the symmetric matrix was then entered into the MDS algorithm. The MDS analysis produced the results shown in Figure 1. The fitting of the data was satisfactory (Kruskal's stress 1= .63; R^2= .62) and the two dimensions representing longitude and latitude can clearly be distinguished. In the interpretation of this figure, one needs to keep in mind that the orientation of the axes is not unique and different orientations can have a constant difference of 180^0.

Figure 1. MDS of LSA cosine matrix of wind directions



Observation 2

If LSA can predict longitude and latitude of words related to wind directions, one could make the prediction it may also be able to predict spatial dimensions of geographical information like place names, states and continents.

When semantic associations were determined between continents like *Europe*, *Asia*, *North America*, *South America*, *Australia* and *Africa*, and entered in an MDS we found a Kruskal's stress 1 = .533 and R^2= .210. The MDS mapping was however very unsatisfactory, with distortions for all continents.

We repeated these analyses using place names of world cities (*London*, *Paris*, *New York*, *Los Angeles*, *San Francisco*, *Amsterdam*, *Moscow*, *Tokyo*, *Beijing*, *Madrid*, *Berlin*, *Washington*, *Chicago*; Kruskal's Stress 1 = .566 and R^2= .582) and largest US cities (*New York*, *Los Angeles*, *Chicago*, *Houston*, *Philadelphia*, *San Diego*, *Detroit*, *Dallas*, *Phoenix*, *San Antonio*, *San Jose*, *Baltimore*, *Indianapolis*, *San Francisco*, *Jacksonville*, *Columbus*, *Milwaukee*, *Memphis*, *Washington*, *Boston*; Stress 1 = .658 and R^2= .330). MDS mapping showed similar distortions as the MDS mapping of continents and were not satisfactory. They are not presented here for paper size limitations.

The analyses mapping semantic associations between proper names did not show the prominent patterns that were found in the first study. At least two explanations can be given for this performance. First, many of the geographical names consisted of two words (e.g. *North America*, *South America*; *San Diego*, *San Francisco*). LSA will have taken the semantic associations between these string pairs and considered them as separate words. New Jersey and New Mexico will therefore be necessarily related because of their sharing of one word. The second explanation for these unsatisfactory findings lies at the heart of the issue: there is simply not enough context for proper names to be meaningful enough. From a cognitive perspective, place names need to be learned. It is hard to induce them from context only. If this explanation has any validity, one would expect that by comparing more frequent spatial words that are used in a variety of context. This is what will be done in Observation 3.

Observation 3

In the third observation we took common lexical items (nouns) that are generally associated with a particular space: *floor*, *ceiling*, *wall*, *door*, *window*, *table*, *chair*, *vase* and *flowers*. Clearly, there is not one spatial orientation of these items (flowers can be under the table and vases next to chairs). It is however not hard to imagine a typical schematic (Schank & Abelson, 1977) representation of this. We computed a matrix of 8 x 8 LSA cosine values and derived a Multidimensional Scaling (MDS) representation. As in the previous observations the mapping was not satisfactory (Stress = .533, R^2=.464). Figure 2 shows the MDS plot of the eight items in a two-dimensional space. The argument can be made that mapping of these items very much relies on subjective reliability. However, Dimension 1 seems to correspond with a vertical axis (with *ceiling* and *floor* being furthest from each other and the other items in between), dimension 2 with a horizontal axis. The mapping of these items thus suggests that common meaningful words can be used to determine spatial dimension. The question is how common and how meaningful? We have compared proper names and common concrete nouns. Would MDS also work with frequent grammatical items like prepositions? This is explored in Observation 4.

Observation 4

We used seven spatial prepositions to compare: *above*, *under*, *left*, *right*, *next*, *in* and *through*. Grammatical items in general and prepositions in particular, are ambiguous in meaning. The word *right* has several meanings (including appropriate, entitlement, but also the antonymous meaning of *left*, which in turn would be a past tense verb or a direction). If it is the case that LSA highly depends on context, we would expect that relations between prepositions would map well onto a two-dimensional space. An MDS representation of the 7 x 7 LSA cosine matrix resulted in a relatively high stress value (.541, R^2 = .609). Nevertheless the two-dimensional space showed the correct patterns (Figure 3): *above* and *under* being orthogonal to ‘left’ and ‘right’. Even for grammatical items mappings are obtained that do not violate our assumptions about the world.

Observation 5

In Observation 2 we found that spatial proper names did not give a satisfactory MDS presentation. For temporal names, like days of the week, the representation was acceptable. One would predict that lexical items would provide an even better representation.

In the final observation we therefore used temporal markers (*yesterday*, *today*, *tomorrow*, *past*, *present* and *future*). Again, the two-dimensional MDS representation yielded a Kruskal’s stress 1 = .424 and R^2 = .293. A graphical representation, presented in Figure 5, shows a past–present–future axis (Dimension 2), and – though less prominent a yesterday–today–tomorrow axis (Dimension 1). Together with the previous observation, this suggests that an MDS representation of the LSA similarity matrix allows for a mapping of words on a temporal dimension.

Figure 2. MDS of LSA cosine matrix of objects in space.

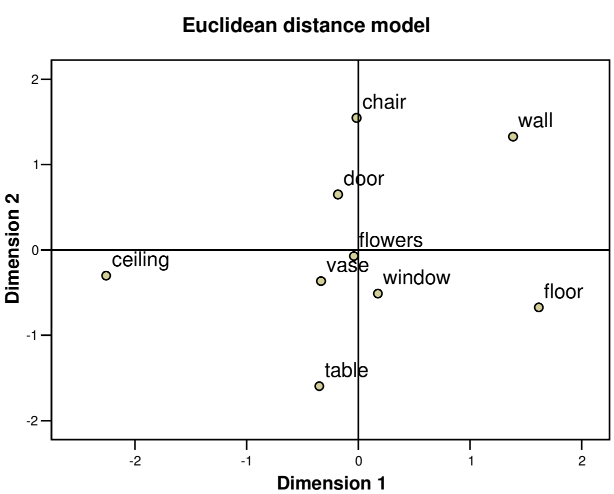


Figure 3. MDS of LSA cosine matrix of prepositions

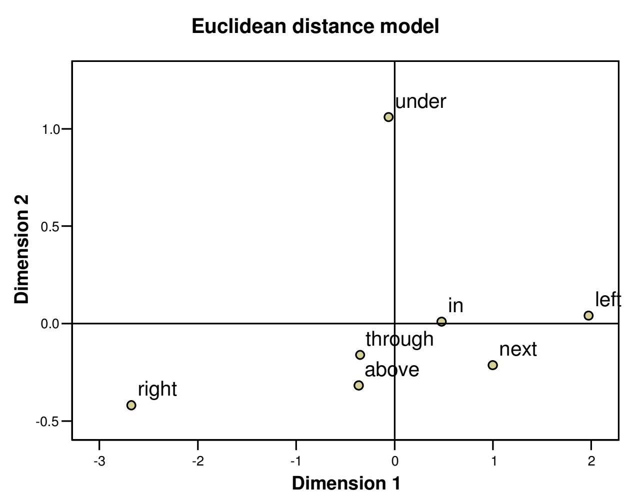
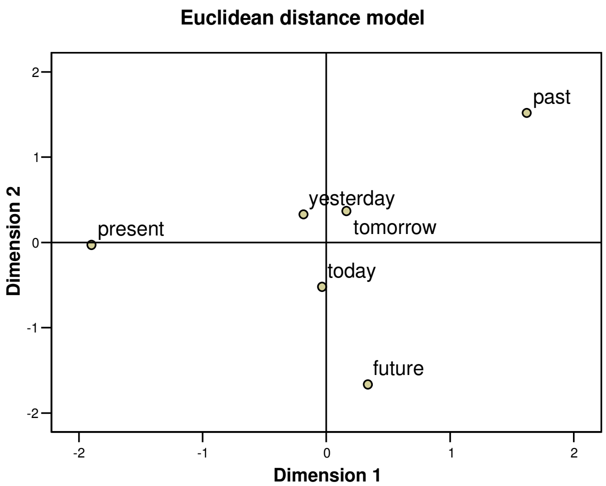


Figure 5. MDS of LSA cosine matrix of temporal markers



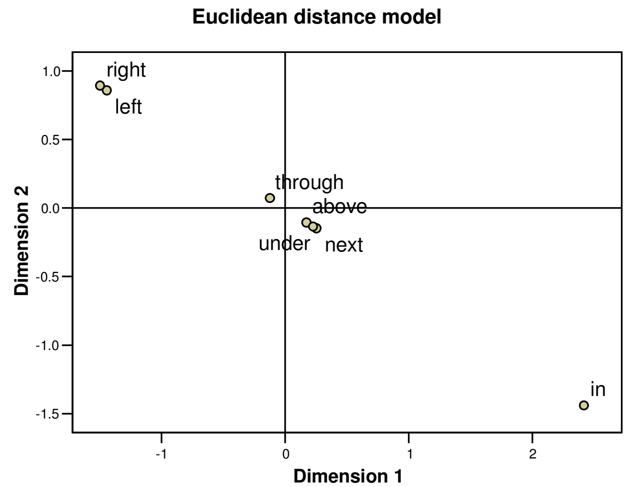
Discussion

NL-KR techniques like LSA have been powerful from an application-driven perspective as well as a cognition-driven perspective. Most cognitive scientists agree that LSA cannot perform optimally on typical embodied dimensions of time and space. Some (Barsalou, 1999; Glenberg & Robertson, 2000) argued that this is due to the fact that LSA is a fundamentally amodal symbolic presentation. Others (Landauer, 1999; Landauer & Dumais, 1997) would argue that this is due to the fact that LSA is deprived of perceptual information. Our preliminary results suggest that LSA in fact performs reasonably well on temporal and spatial dimensions. The current exploratory results may yield at least two immediate questions: 1) Why is it that researchers have not considered this wealth of information before? 2) How can we explain finding temporal and spatial dimensions from (higher-level) co-occurrences of words and contexts?

The answer to the first question lies in the comparison techniques that have been used. In most, if not all, LSA studies pairs of cosine values are compared. The strength of LSA, however, lies in comparing one item in relation to all other items. In this study we have for instance not compared the LSA cosine between the words *North* and *South* only, but instead compared the relationship between *North* and *South* in the context of all other relevant relationships. Our results suggest that an MDS representation of semantically related words provide some advantages in representing knowledge.

So what is the explanation behind these findings? We are currently exploring the answers. For instance, will co-occurrences counts between terms produce the same degree of interpretability?

Figure 6. MDS of prepositions bigram frequency



We took the prepositions used in Observation 4 and computed their bigram counts in the TASA corpus. We then supplied these frequencies to the ALSCAL algorithm. Despite an almost perfect fitting of the data (Stress = .08, $R^2 = .99$) the scaling is far less intuitive, as shown in Figure 6.

So is there anything in the LSA space that identifies the spatial information? Berry & Fiero (1996) explored possible geometric representations with LSA spaces and tried to determine if certain properties can be observed by simple projection, i.e., by projecting 300 dimensional LSA space onto a two dimensional subspace determined by the first two dimensions. Accordingly, we calculated all combinations of pairs of the 300 LSA dimensions for the words *North*, *South*, *West* and *East* (Observation 1). If we can identify two dimensions that mimic the MDS longitude and latitude dimensions, we may conclude the answer lies in LSA. Alternatively, the answer must lie outside LSA. For the four wind directions four 300 dimensional vectors were obtained from the LSA TASA space. The total number of pairs of dimensions is thus 300 x 299 / 2 = 44850. For each pair of dimensions (i, j), we got the coordinates of the four vectors on these two dimensions: (north(i), north(j)), (south(i), south(j)), (east(i), east(j)) and (west(i), west(j)). They can be seen as the 'images' of the four words in the selected two-dimensional space. We then connected the images of 'south' to 'north', and 'east' to 'west' to form two lines. The angle (ranging from 0 to 90 degrees) was then calculated. If the dimension pairs tend to pick up the spatial information, we would expect that the angles should be close to 90. Results however showed that the angles were uniformly distributed in the interval [0, 90]. This shows that what we have observed is not some linear property of LSA space (simple projection of multi-dimensional space onto some two-dimensional subspace). Instead, it is the MDS that extracts some special properties of the LSA space that is not directly observable with the

LSA framework. It thus appears that the patterns apparent in the MDS solutions are of a higher-order nature. The exact nature is currently investigated.

NL-KR systems each have their own strengths and weaknesses. One of the strengths of LSA is that it has proven to be powerful in a variety of applications. Another is that it could simulate aspects of human cognition. Its weakness is that it is presumably not good at dealing with dimensions of space and time in applications. Another is that it fundamentally cannot simulate human cognition because it is not embodied in space and time. Our results raise doubts about these weaknesses. Space and time seem to be represented in LSA if a higher-order technique can make these temporal and spatial dimensions explicit. This suggests that an amodal symbolic knowledge representation like LSA is powerful because it can be used in a variety of applications and also simulates many aspects of human cognition including some embodied concepts.

Acknowledgments

This research was supported by NSF (IIS-0416128; REC-0106965). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavior and Brain Sciences, 22*, 577-660.

Berry, M.W. & Fiero, R.D. (1996). Low rank orthogonal decompositions for information retrieval applications. *Numerical Linear Algebra with Applications*, *l*, 1-27.

Glenberg, A. M. & Robertson, D. A. (2000) Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379-401.

Graesser, A.C., McNamara, D.S., Louwerse, M.M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, *36*, 193-202.

Graesser, A., VanLehn, K., Rose, C., Jordan. P., and Harter, D. (2001). Intelligent tutoring systems with coversational dialogue. *AI Magazine, 22*, 39-51.

Graesser, A.C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., & Person, N., & the TRG (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments, 8*, 128-148.

Hu, X., Cai, Z., Franceschetti, D., Penumatsa, P., Graesser, A.C., Louwerse, M.M., McNamara, D.S., & TRG (2003). LSA: The first dimension and dimensional weighting. *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 587-592). Boston, MA: Cognitive Science Society.

Kintsch, W. (1998) *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kruskal, JB, and Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage University Series.

Landauer, T. K. (1999). Latent Semantic Analysis (LSA), a disembodied learning machine, acquires human word meaning vicariously from language alone. Commentary on Barsalou (1999). *Behavior and Brain Sciences, 22*, 624-625.

Landauer (2002). On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The psychology of learning and motivation, 41*, 43-84.

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.

Louwerse, M.M., Bard, E.G., Steedman, M., Hu, X., Graesser, A.C. (2004). *Tracking multimodal communication in humans and agents*. Technical report, Institute for Intelligent Systems, University of Memphis, Memphis, TN.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variations in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 843-848). Mahwah , NJ : Erlbaum.

Louwerse, M.M. & Ventura, M. (in press). How children learn the meaning of words and how computers do it (too). *Journal of the Learning Sciences*.

Miller, G. A. (Ed.). (1990). WordNet: An on-line lexical database. *International Journal of Lexicography 3*, 235-312.

Pasca, M. & Harabagiu, S. (2001). The informative role of WordNet in open-domain question answering *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations* (138-143). Carnegie Mellon University, Pittsburgh PA.

Rus, V. & Moldovan, D.I. (2002). High precision logic form transformation. *International Journal on Artificial Intelligence Tools, 11,* 437-454.

Schank, R. C. & Abelson, R. P. (1977) *Scripts, plans, goals and understanding: an inquiry intohuman knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.