International Journal on Artificial Intelligence Tools
Vol. XX, No. X (2005) 1–20

COGNITIVELY INSPIRED NLP-BASED KNOWLEDGE REPRESENTATIONS: FURTHER EXPLORATIONS OF LATENT SEMANTIC ANALYSIS

MAX LOUWERSE *

ZHIQIANG CAI

XIANGEN HU

MATTHEW VENTURA

PATRICK JEUNIAUX

Department of Psychology / Institute for Intelligent Systems,
Memphis, Tennessee 38152, United States of America
{mlouwerse, zcai, xhu, mventura, pjeuniax} @memphis.edu

Received (30 June 2005)
Revised (27 September 2005)
Accepted (23 November 2005)

Natural-language based knowledge representations borrow their expressiveness from the semantics of language. One such knowledge representation technique is Latent semantic analysis (LSA), a statistical, corpus-based method for representing knowledge. It has been successfully used in a variety of applications including intelligent tutoring systems, essay grading and coherence metrics. The advantage of LSA is that it is efficient in representing world knowledge without the need for manual coding of relations and that it has in fact been considered to simulate aspects of human knowledge representation. An overview of LSA applications will be given, followed by some further explorations of the use of LSA. These explorations focus on the idea that the power of LSA can be amplified by considering semantic fields of text units instead of pairs of text units. Examples are given for semantic networks, category membership, typicality, spatiality and temporality, showing new evidence for LSA as a mechanism for knowledge representation. The results of such tests show that while the mechanism behind LSA is unique, it is flexible enough to replicate results in different corpora and languages.

Keywords: Knowledge representation, Latent Semantic Analysis, semantic similarity.

1. Introduction

The questions how humans represent knowledge and how computers can do this have been shown to be intrinsically related. In one of the earliest semantic network representations, Quillian[1] proposed a model for semantic knowledge in the form of the computational model TLC (Teachable Language Comprehender). His model explored the way word and concept knowledge could be stored in a computer program that

* Max M. Louwerse, Department of Psychology / Institute for Intelligent Systems, University of Memphis, 202 Psychology Building, Memphis TN 38152.

represented a model for language comprehension. According to Quillian computers – like humans – can store categorical information in a network where properties like *yellow* are assigned to concepts like *canary* but not to concepts like *robin*, but properties *can fly* can be assigned to both, as can properties like *breathes*, which in turn are also assigned to concepts like *shark*. Features like *is-a*, *has-a*, *is-able-to* assigned to these nodes of the hierarchy allow for a simple knowledge representation system. Collins and Quillian[2] found that such a hierarchical semantic network has psychological implications: sentence verification times for nodes that are close to each other (*terrier is-a dog*) are faster than nodes that are further away from each other (*terrier is-a mammal*). One of the problems with Quillian's model was that we are generally faster in deciding that a canary is a bird than that an ostrich is a bird, and are likely using different properties when we compare a canary to a shark (e.g. the property *yellow* is likely not even activated). This led Smith, Shoben and Rips'[3] proposal of a feature-based model that used features that are *defining* for a concept and therefore always present in its instances (*have-wings* for *birds*) and features that are *characteristic*, somewhat typical of a concept but not present in each instance (*can-fly* for *birds*). Their experiments showed that a comparison of defining and characteristic features between concepts determines its retrieval of information. In other words, semantic relatedness rather than position in a logical hierarchy was the determining feature in sentence verification. The problem with these notions of semantic networks is that they were still rather rigid. More sophisticated structures have been proposed since, including those of frames[4], schemas[5] and scripts[6]. These various proposals have made suggestions how human memory is organized both semantically and schematically. The principal difference between schema theory[4,5,6] and semantic networks[1,2,3] is that the former represent these with molecular phenomena (fixed properties, interrelated to each other), whereas the latter deals with molar phenomena (events, spatial information).

The semantic network and frame theory proposals of knowledge representation formed the origin of complex computational implementations. As Gaines[7] pointed out the visual representation of the initial knowledge structures was attractive, but these models had problems of imprecise semantics. Various knowledge representation systems have been inspired by semantic networks and schema theories. FRL (Frame Representation Language) and KRL (Knowledge Representation Language)[8,9], KRYPTON[10], BACK[11], LOOM[12], CLASSIC[13], KRIS[14] and CYC[15] are some of examples of knowledge representation systems. Despite their benefits for a range of computer science applications, the problem with this kind of knowledge representation system has been the same as the problem with Quillian's semantic network: they don't seem to exploit the complexity of natural language.[16].

The question of course is why one wants to base knowledge representation systems on natural language in the first place. Natural language is generally not systematic. It is highly ambiguous both lexically and syntactically. The word "bear" can be a noun or a verb, the verb having the meaning *give birth to* or *put up with*. Words are often polysemous, that is, they often have one or more related meanings (e.g. "foot" in *foot of a human* or *foot of a mountain*). Or they are semantically associated, but in an opposite sense ("small" vs. "large", "up" vs. "down"). But the problems in systematicity go beyond the word level. Even if we disregard phrases and sentences (consider the three meanings of "The teacher strikes idle kids"), the relations between the concepts are as problematic as the meaning properties of the concepts. Words can be synonymous ("kid" and "child"), but can also share aspects of their meaning ("bachelor", "father", "uncle", "husband" all share *male* and *human*).

Despite the fact that natural language itself may not be all that systematic, natural-language based knowledge representations have a number of advantages over knowledge representation systems that are not. Firstly, they tend to be intuitive in their construction and integration with a intuitive semantics. Secondly, no new classification system needs to be developed, since there is already an ontology that underlies language itself. Thirdly, natural language ontologies consist of categories that we can relate to: representations are relatively easy to share. Fourthly, most knowledge representations are used for natural language domains. It therefore seems efficient to eliminate a translation step into some form of formal language. Finally, natural language is expressive. Doing justice to this expressiveness requires a knowledge representation in natural language. In sum, there are plenty of reasons why natural-language based knowledge representations (NL-KR) are beneficial.

A two-tier distinction can be made between NL-KR systems, ones that are application-driven and ones that are cognition-driven. In the former few similarities can be found between the knowledge representation system and the knowledge representation in humans in terms of architecture or processes. This category of systems primarily serves the need to be applied. Most NL-KR systems are application-driven: they provide the necessary tools for the computational linguist and are precise in nature. An example of such an application-driven system is WordNet[17], 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. For instance, Wordnet has been used for logic form representations[18] and question-answering systems[19]. It has also been used in applications measuring the difficulty of text[20].

One could of course immediately question the fact that WordNet is based on cognitive decisions regarding the relationships between words. However, for the reasons outlined earlier[16] it is difficult to consider it to be a cognitive model. As with dictionaries, WordNet provides an excellent language resource, but it can say little how humans acquire, use or process language. For instance, efforts to measure semantic similarity using WordNet have so far not proven to be successful[21,22].

The other NL-KR system approach is cognition-driven. These systems simulate – or claim to simulate – aspects of human cognition. Examples of such cognition-based systems are Hyperspace Analog to Language (HAL) model[23] and Latent Semantic Analysis (LSA)[24]. As shown in the brief historical overview presented earlier there has always been a strong connection between knowledge-based representations proposed in psychology and those modeled in computer science, suggesting that psychological models can contribute to computer models and vice versa. Often cognition-driven knowledge representation systems tend to be less precise in terms of systematicity than the application-driven systems. This is not a surprise: human cognition is not as rigid as a computational model may want to be. In other words, it seems that cognition could be more easily modeled with heuristics than with algorithms. Despite their lower precision, cognition-based NL-KRs however have some important advantages. Firstly, this lower precision can be viewed as an advantage in that it allows the representation of approximate relations between words (i.e. what human language seems to be). A word can be *more or less* synonymous or can share a relation with another word on one (even *unspecified*) dimension. Secondly, they are generally tested experimentally. That is, psychological data is compared with the system. Thirdly, they can be related to theoretical cognitive frameworks. Both HAL and LSA use large corpora to develop cognitively plausible high-dimensional semantic models that mimic aspects of cognitive phenomena by capitalizing on the context in which words and sentences appear. LSA represents the corpora as a high-dimensional co-occurrence matrix of words in texts, and reduces its dimensions using singular value decomposition. HAL builds a semantic word co-occurrence matrix, which is weighted according to co-occurrence frequency. Because HAL is less commonly used and because it is less readably available, we will focus here on LSA.

The current paper will discuss the many aspects of LSA as a NL-KR system. The paper is divided into two parts. In the first part (sections 2 and 3) we discuss the mathematics behind LSA and the many applications of LSA, including essay grading, coherence metrics and intelligent tutoring. In the second part (sections 4, 5, 6 and 7), we hypothesize how LSA can be seen as a model of human symbolic processing and show a number of experiments that provide a new outlook to LSA and natural-language based knowledge representations.

2. 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 (where -1 indicates a minimal semantic relation and 1 indicates a maximal semantic relation) 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 $A = \left(f_{ij} \times G(i) \times L(i, j)\right)_{m\times n}$ 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 $G(i)$ is the global weighting for term i. 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 smaller 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 rows of 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 vector of k dimensions. The semantic relationship between words can be estimated by taking the normalized dot product (cosine) between two vectors[25, 26].

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[24]. An example of the semantic similarity based on co-occurrences is presented in Table 1 and Figure 1. Note that "canary" and "robin " never occur in the same document, but because of the higher-order relationship with words like "swim" they are semantically associated , as shown in dimension 2 of Figure 1.

Table 1. Term-by-document matrix A

	Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	Paragraph 5
Canary	1	0	0	0	0
Robin	0	1	0	0	0
Swim	1	1	0	1	1
Fly	0	0	0	1	0

3. Latent Semantic Analysis Applications

Since 1990, LSA has been used in a variety of applications. Initially it was used for information retrieval purposes[27,28], where it scored an average precision up to 30% in standard accuracy measures better than other methods. Dumais[29] reported that LSA was 16% better than other standard keyword vector methods. Interestingly, LSA performed best in cases where the queries and retrieved documents shared only a small number of words and at high levels of recall. Apparently, LSA is able to go beyond key word matching and capture synonymy.

To test LSA's knowledge of synonyms Landauer and Dumais[24] tested how well LSA would pass the Test of English as a Foreign Language (TOEFL) test by the Educational Testing Service that every foreigner at an American University needs to take. On 80 multiple choice test items an LSA space of approximately 2000 pages scored 64% correct, compared with 33% correct for word-overlap methods, and 64% correct for the average student taking the test.

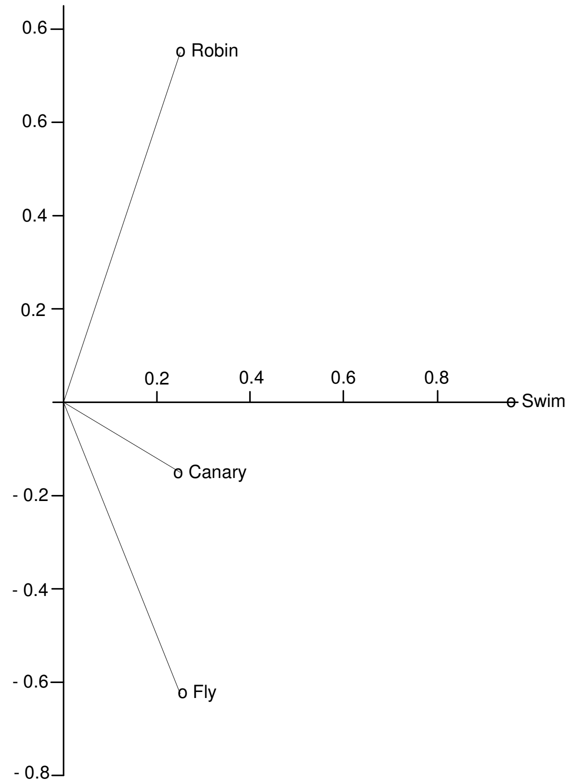

Fig. 1. Example of dimensionality reduction using SVD.

But LSA can do more than identifying synonyms in language. Landauer, Foltz and Laham[30] trained LSA on introductory psychology textbooks and tested how well it would identify the right answer in multiple choice questions given by the textbook publishers. Performance was not outstanding, but compared to class average, LSA would receive a passing grade. When Landauer *et al*. created on LSA space of textbook and student essays they found that LSA correlated better with expert graders than these graders did with each other. The success of LSA in grading essays has led to the Intelligent Essay Assessor (IEA), now a commercial product that is commonly used. The IEA computes content, style, and mechanics of an essay as well as self validation, confidence and counterfeiting measures.

LSA is also the engine behind Summary Street, a reading comprehension and writing instruction tool. Students write summaries of a text and Summary Street evaluates these summaries by comparing them to the text and providing feedback about the summary content and writing mechanics[31].

4. LSA as Knowledge Representation Theory

Within the learning environment LSA has even proven to be successful as the model of a long term memory of a conversational partner. The intelligent tutoring system AutoTutor[32] 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[33,34].

Foltz, Kintsch and Landauer[35] compared how well LSA could identify changes in content within a text. They reanalyzed the texts in two studies that manipulated the coherence of texts and assessed readers' comprehension and found that LSA can measure the coherence in text adequately. LSA as coherence metrics also plays an important role in Coh-Metrix[20], 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.

Kintsch[36] used LSA to design a computational model of metaphor comprehension by computationally modeling the interaction between the meaning of the topic and vehicle terms of the metaphor. In a prediction algorithm that combines Kintsch's Construction Integration model with LSA the metaphor ("my lawyer is a shark") is understood as merging the topic vector ("lawyer") with selected features of the vehicle vector ("shark"). Kintsch[37] also applied LSA to identify the theme and subthemes of a text. By comparing the propositions of a text LSA was able to construct the macrostructure, and thereby the theme, of a text[38].

Louwerse[39] applied LSA to literary texts to determine the style of the author (idiolect) and groups of authors based on gender and literary period (sociolect). Where simple keyword algorithms failed, LSA was able to classify texts in terms of idiolect and sociolect on the basis of lexical consistency.

Finally, Louwerse, *et al.*[40] applied LSA in combination with other computational linguistic measures to a set of corpora used by Biber[41] to determine variations in written and spoken registers. Louwerse *et al.* were able to replicate most of the textual dimensions that Biber found in his work. LSA helped to distinguish speech from writing, factual information from situational information, topic consistency versus topic variation, elaborative versus constrained, and narrative versus non-narrative.

Most of these applications using LSA are text-based. Quesada, *et al.*[42] have however shown that LSA's basic approach can also be applied to any other knowledge domain with high numbers of weak relations between tokens. They used LSA in Complex Problem Solving where the tokens represented states of the system or the actions to control the problem.

In sum, LSA has been widely used in a variety of applications, ranging from text analysis, metaphor comprehension to problem solving. In most of these cases the application simulates what a human would do otherwise. Indeed, Landauer and Dumais[24] have argued that LSA is a theory of human acquisition, induction and representation of knowledge.

Since its early days LSA has been considered to provide a solution to Plato's problem, the psychological problem of how humans observing a relatively small set of events can construct knowledge representations that are adaptive in a large, potentially infinite variety of situations[1]. LSA does this by mapping initially meaningless words into a continuous high dimensional semantic space, more or less simulating cognition[43]. More specifically, it uses a first-order process that associates stimuli (words) and the contexts they occur in (documents). This process is very much like psychological 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[24,44,45].

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[46,47] 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[43,44]. 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"[46].

Louwerse[48] provides a different response, arguing that LSA does not have to include perceptual information for it to model large aspects of language processing. Based on Deacon[49] he proposed a theory of symbol interdependency that explains how amodal symbol systems become transduced from perceptual states. Deacon introduced a model of symbolic transduction/reference (read symbol interdependency), which is loosely based on Peirce's[50] distinction of signs into icons, indices and symbols. Icons are mediated by a similarity between the sign and the object it refers to, like a photograph resembling a person. Indices are mediated by contiguity between sign and object, like a footprint indicating the presence of a deer. Symbols are mediated by an arbitrary, conventional relationship between sign and object, like a heart-shape symbolizes love. Different levels of interpretation can now be formed based on the hierarchical relationships of this classification, whereby symbolic relationships are composed of indexical relationships, and these indexical relationships are composed of iconic relationships. Symbols thereby refer to indices and ultimately icons, but also have interrelations between themselves: "The correspondence between words and objects is a secondary relationship, subordinate to a web of associative relationships of a quite different sort, which even allows us to reference to impossible things"[51]. Symbolic interdependency provides an explanation why humans have language and animals do not, how children gradually learn the meaning of words and how the evolution of language has come about. Psychologically, symbol interdependency has the advantage of being mnemonic support. We can more easily retrieve words based on the interdependency of these words with other words, and we can strengthen the (indirect) links between symbols and objects by means of other symbols. A consequence of symbol interdependency is that language is likely to be structured in a way that facilitates retrieving embodied structures[48]. That is, perception and cognition structure language which means that symbols can be (and are occasionally) grounded in perceptual experiences, but do not always have to be because symbols can retrieve meaning through other symbols. It is this symbolic level that LSA seems to model.

If the theory of symbol interdependency is valid, and if LSA is a model of human knowledge representation, then there are a number of psychological data LSA will have to account for. We will next explore some of the unexplored data.

5. Typicality measures

This paper started out with Quillian's[1] theories regarding semantic networks. One of the first questions to be answered is whether LSA can approximate these networks. Collins & Quillian's[2] frequently quoted hypothetical memory structure serves as a good example. The problem with a comparison of the words in the structure using LSA is that LSA will semantically associate n-grams that share words ("has fins", "has wings") because of the presence of the identical words ("has"). To put all nodes on an equal playing field, we selected one word that is typical for a node (e.g. "upstream" instead of "swims upstream to lay eggs"). As we outlined earlier, LSA's knowledge of the world comes from a set of corpora segmented into smaller chunks of context (documents) like paragraphs or sentences. We 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. An LSA space was created for the TASA corpus.

To do justice to the hierarchical nature of the semantic network, the cosine matrix was analyzed using a hierarchical cluster analysis. The matrix of 26 x 26 cosine values was used in a Hierarchical Cluster Analysis, employing the between-groups linkage algorithm on a Euclidean distance measure. The cluster analysis then forms pairs of variables which are close on one dimension (or combinations of dimensions) and forms clusters of those pairs and forms higher-order combinations until all variables are included in the cluster. Results of the hierarchical clustering are presented in Figure 2. Features generally did not cluster with the concept, but rather with other similar features (e.g. *yellow* and *pink*, instead of *yellow* and *canary* or *pink* and *salmon*). Nevertheless, features like *wings*, *feathers* and *fly*, key for the concept *bird* did cluster, and so did *fin*, *gills* and *swim* cluster with *fish*. From the network we also learn that *animals eat*, have *skin*, *move*, *bite*, are *dangerous* and *edible*.

It is somewhat difficult to match the presented hierarchy to Collins & Quillian's[58] hierarchy, because of the clustering of features. The analysis was therefore repeated using just the main nodes of the network (*animal*, *bird*, *canary*, *ostrich*, *fish*, *salmon* and *shark*). A hierarchy emerged as presented in Figure 3, that very much matches to Collins & Quillian[2] theoretical proposal.

The problem with analyses like these is that they are very exploratory and can capitalize on chance. What is a good match for one person, may be an unsatisfactory match for another. We therefore conducted more conventional statistical studies on a larger data set of psychological data. Rosch[52] conducted a number of studies investigating category membership in which participants rated the typicality of members of a category on a scale, showing that participants were consistent in their ratings (e.g. *robin* is a more typical member of the category *bird* than is *chicken*). Furthermore, participants were faster in judging whether a picture belongs to a certain category when the picture showed a typical member than when it was a non-typical member. Eight categories (*fruit*, *science*, *sport*, *bird*, *vehicle*, *crime*, *disease* and *vegetable*) with six members in each category were taken from Rosch's work[52,53].

Results are presented in Table 2. Overall, there was a significant negative correlation between rank of category and cosine value, whereby high rank (e.g. 1) corresponds to a high cosine value (e.g. $cos = .64$) and a low rank (e.g. 6.2) to a low cosine value (e.g. $cos = .23$) (*Spearman* $r(47) = -.618$, $p < .001$; Note that the word "embezzling" did not occur in the corpus, so no cosine value could be computed). Correlations per category showed that five out of the eight categories (*bird*, *crime*, *fruit*, *sport*, *vegetable*) had significant correlations. The remaining three categories (*disease*, *science*, *vegetable*) did have the expected pattern, though these did not receive the significance level, most likely due to the small number of cases (six per category). These findings suggest that LSA can simulate semantic networks as well as psychological data coming from these semantic networks.

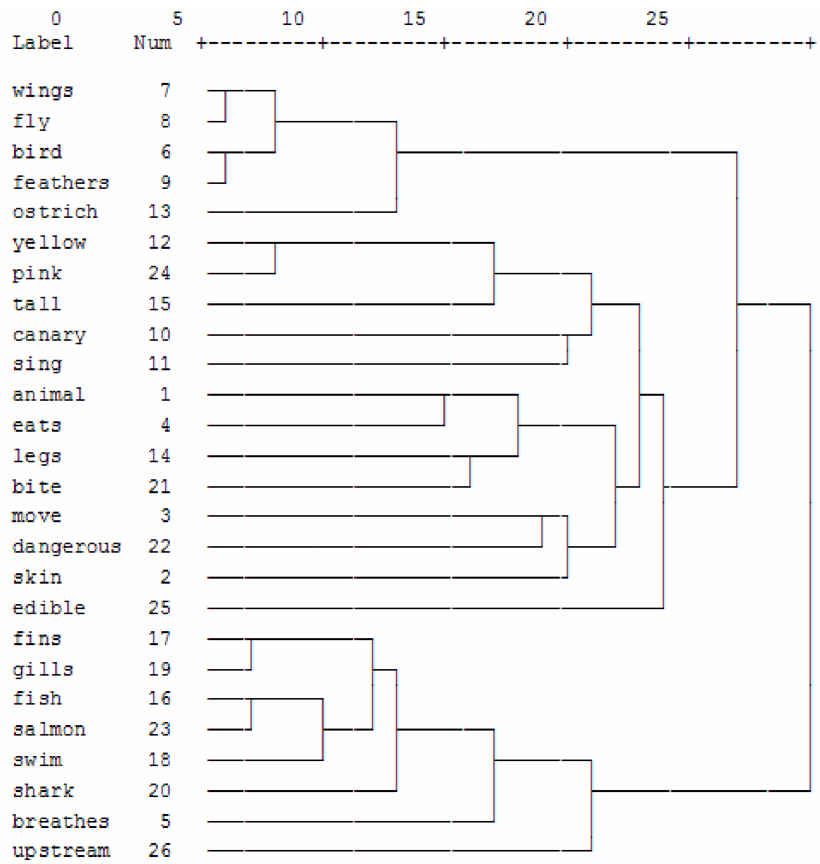

Fig. 2. Complete overview LSA hierarchy of Collins & Quillian[2]

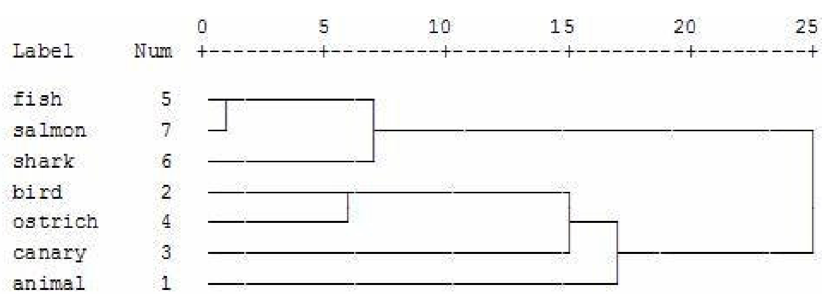

Fig. 3. Selected overview LSA hierarchy of Collins & Quillian[2]

Table 2. Typicality rank from Rosch[52] and LSA cosine values

Category	Member	Prototypicality	LSA cosine
fruit	apple	1.3	0.47
	plum	2.3	0.27
	pineapple	2.3	0.38
	strawberry	2.3	0.33
	fig	4.7	0.02
	olive	6.2	0.23
science	chemistry	1	0.64
	botany	1.7	0.66
	geology	2.6	0.71
	sociology	4.6	0.44
	anatomy	1.7	0.41
	history	5.9	0.24
sport	football	1.2	0.76
	hockey	1.8	0.75
	wrestling	3	0.48
	archery	3.9	0.24
	gymnastics	2.6	0.6
	weight lifting	4.7	0.09
bird	robin	1.1	0.52
	eagle	1.2	0.8
	wren	1.4	0.4
	chicken	3.8	0.31
	ostrich	3.3	0.57
	bat	5.8	0.2
vehicle	car	1	0.47
	boat	2.7	0.04
	scooter	2.5	0.12
	tricycle	3.5	0.01
	horse	5.9	0
	skis	5.7	0.05
crime	murder	1	0.75
	assault	1.4	0.41
	stealing	1.3	0.41
	embezzling	1.8	n/a
	blackmail	1.7	0.14
	vagrancy	5.3	0.25
disease	cancer	1.2	0.43
	measles	2.8	0.87
	cold	4.7	0.14
	malaria	1.4	0.68
	muscular	1.9	0.16
	rheumatism	3.5	0.11
vegetable	carrot	1.1	0.28
	asparagus	1.3	0.42
	celery	1.7	0.45
	onion	2.7	0.25
	parsley	3.8	0.39
	pickle	4.4	0.26

Rosch's[52,53] conducted reaction time experiments that allowed for a distinction between central members of a category (*doll* in the category *toy*) and peripheral members (*skates* in the category *toy*). We next explored whether LSA could match the categorization findings. Twenty-four central and twenty-four peripheral members of the twelve categories (*toy*, *bird*, *fruit*, *sickness*, *relative*, *metal*, *crime*, *sport*, *vehicle*, *science*, *vegetable*, *body parts*) were taken and cosine values between member and category were computed. If LSA is able to measure category membership, it should find a significant difference between the category-central member cosine values and the category-peripheral member cosine values, with higher scores for the central than for the peripheral scores. A paired-samples *t*-test confirmed this hypothesis (t (1, 23) = 4.59, $p <$ .001) with higher scores for the category vs. central members (M = .46, SD = .22) than the category vs. peripheral members (M = .26, SD = .14).

The above results suggest that LSA is able to represent concepts and features in a semantic network fashion in the line of what Quillian[1] proposed and what Rosch[52,53] found experimentally. As was shown in the introduction of this paper, we know since Quillian's work that human memory is organized both semantically and schematically. We have shown that LSA can capture these networks, category membership relations and typicality judgments[54]. These analyses have compared the semantic of a word with the semantics of another word, much in the line of the strengths of LSA outlined in the description of its applications. A real challenge in terms of semantics would be to capture relationships that are more abstract, like temporality and spatiality.

6. Temporality measures

Louwerse, *et al.*[55] made the argument that in making semantic judgments, humans generally consider semantic fields of words rather than pairs of words. That is, *shark* and *salmon* may be very similar when compared with an *elephant*, but turn out to be very different when comparing with a concept like *dangerous*. 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 order to determine latent relations between items. The question whether LSA can represent temporality and spatiality lies in the line of Louwerse *et al.*'s argument.

Take the cosines values for days of the week using the TASA LSA space. When the cosine is computed between Monday and the other days of the week, the highest cosine values are found for Tuesday (*cos* = .78), Thursday (*cos* =.74) and Friday (*cos* =.75). But the argument can be made that humans know the meaning of *Monday* not just by its relation with another day of the week, but by the interrelations of all days of the week. Louwerse, *et al.*[55] argued that the strength of LSA lies in comparing one item in relation to all other items. That is, rather than comparing the LSA cosine between the words "Monday" and "Tuesday" only, the relationships should be considered between "Monday" and "Tuesday" in the context of all other relevant relationships. A Multidimensional Scaling representation of semantically related words thereby provides some advantages in representing knowledge.

The LSA matrix of cosine values was supplied to an ALSCAL algorithm to derive a Multidimensional Scaling (MDS) representation of the stimuli[56]. That is, the matrix of LSA cosine values is transformed into a matrix of Euclidean distances and these distances are scaled multidimensionally by comparing it with arbitrary coordinates in an *n*-dimensional space (low cosine values correlates with large distances, high values with short distances). The coordinates are iteratively adjusted such that the Kruskall's stress is minimized and the degree of correspondence maximized. The fitting of the data was satisfactory (*Kruskal's stress 1* = .209; R^2= .897) with a one-dimensional scaling. When the stimulus coordinates were rank ordered and compared to the actual rank order of the days of the week (Monday = 1, Tuesday = 2, … Sunday = 7) the significant correlation (*Spearman rho* (7) = .82, p = .023) showed that LSA correctly represents the order of the days of the week. A one-to-one comparison between names does not reveal this information.

The same analysis was conducted using the months of the year. A matrix of 12 x 12 LSA cosine values was computed and an MDS representation was derived (*Kruskal's stress 1* = .236, R^2=.980). As with the days of the week, the months were rank ordered (January = 1, February = 2, March = 3 *etc.*) and compared with the rank order obtained from the MDS one dimensional scaling. Again, the correlation was significant (*Spearman rho* (12) = -.580, p = .048) with LSA representing the order of the months of the year correctly. It may be relevant to note here that MDS assigns coordinates arbitrarily. This means that a variable may be placed high on a positive or negative coordinate arbitrarily, resulting in either significant positive or negative correlations.

In addition to measure order conventional times units like days of the week or months of the year, we next investigated whether LSA was able to determine time distances of relative time units ("second", "minute", "hour", "day", "week", "month" and "year") were taken. To add to the temporality these were combined with the words "ago" and "later" (e.g. "a second ago", "a second later"). As before, a 14 x 14 cosine matrix was supplied to an ALSCAL algorithm and an acceptable MDS representation (*Kruskal's stress 1* = .343, R^2 =.821) was derived. Rank ordering the 14 items one a time line ("a year ago … a year later") resulted in a significant correlation with the MDS output (*Spearman rho* (14) = -.835, p < .001), such that "ago" and "later" were plotted away from each other on one axis and "second", "minute", "hour", "day", "week", "month" and "year" (in that order) on the other.

LSA thus seems to be able to temporally order items (days of the week, months of the year, units of time) and capture semantic information from which notions of temporality can be inferred.

7. Spatiality measures

A final test to determine whether LSA captures world knowledge consists of predicting the distances between spatial coordinates. That is, cosine values should be higher in comparing "New York" to "Boston", than for "New York" to "Seattle". Such a hypothesis is not as outlandish as it may sound. It is plausible that in natural discourse those place names that are close in their vicinity occur more frequently in the same documents and, similarly, their contexts will be shared across documents. To test this hypothesis, 28 of the largest cities in the United States were selected and their distances in miles were determined (source: National Geodetic Survey), resulting in a matrix of 28 x 28 = 284 cells. The same 28 cities were entered in LSA using the TASA LSA space and cosines for the 28 x 28 matrix were determined. If LSA can capture actual distances, the prediction would be that there is a negative correlation between distances (lower miles mean closer distances) and cosine values (higher cosine values mean a closer relationship). As predicted a significant negative correlation was found between the distances in miles and similarities in cosine values (r (784) = -.313, $p < .001$).

These results may be explained by the nature of the TASA corpus. It may be the case that textbooks provide lists of place names, though an analysis of the corpus did not provide any evidence for this. Nevertheless, to avoid a corpus bias, a subsequent analysis was run on the Encyclopedia LSA space. This space consists of text from about 30,475 encyclopedia articles, with 60,768 unique terms. Results strongly correlated with the TASA cosine findings (r (784) = .84, $p < .001$) and were slightly higher between the distances in miles with the cosine scores for the encyclopedia (r (784) = -.344, $p < .001$), again suggesting that LSA can capture actual distances from higher-order relationships between co-occurrences.

Of course, the argument could be made that this may work well for cities in the United States because these corpora all contain American texts. It may not be a complete surprise that when Denver is mentioned in a corpus, the likelihood that Dallas is mentioned is high. The question could be raised whether the prediction that LSA can simulate spatial distances works equally well at a more global level, for instance for world cities. Distances were determined for the largest cities in the world (Source: Encyclopaedia Brittanica). These distances in miles were next compared with the cosine values obtained for the 284 comparisons in the TASA LSA space. Again the prediction is that there is a negative correlation between distances in miles and similarities in cosine values. Again, such a significant negative correlation was obtained (r (784) = -.400, $p < .001$). When the same analysis was replicated using the Encyclopedia space, the correlation was slightly higher (r (784) = -.441, $p < .001$), while the correlation between the TASA and Encyclopedia findings was high (r (784) = .960, $p < .001$).

Clearly, the argument made here is not that LSA can be used for map directions. In fact, when an MDS ALSCAL algorithm was run on the data, plots emerged that were far from satisfactory, with a 5-dimensional solution reaching *Kruskal's stress 1* = .103, R^2 = .887 for the TASA space[55]. Ambiguous words (e.g. "Phoenix", "Buffalo", "Washington") and multi-word names ("New York", "New Orleans") may explain why results might be lower than expected. But a more straightforward explanation is that LSA does not measure distances, or rather, measures more than distances. "Chicago" and "New Orleans" may not be close in space, but are close in jazz music. It is therefore even more remarkable that a statistical metrics can capture spatial distances through higher-order relationships in documents only.

The option that these findings can be attributed to a specific corpus have been ruled out by replicating the results in the Encyclopedia LSA space. But perhaps the results can be explained by another factor, for instance the language used. Both the US cities and the world cities were entered in a French LSA space, after being translated. This space consists of 12 months of text from the French newspaper Le Monde. For the US cities a significant correlation was found between the distances and the LSA cosine values (r (729) = -.246, p < .001), while the results strongly correlated with the TASA and Encyclopedia findings (r (729) = .620, p < .001 and r (729) = .688, p < .001 respectively) (Note that the word "Louisville" was not found in the French space, resulting in a lower N). For the world cities, results were equally promising. The correlation with the actual distances was significant (r (676) -.407, p < .001) and correlations with the TASA and Encyclopedia scores (r (676) = .763, p < .001 and r (676) = .783, p < .001 respectively) (The French words "Manilla" (Manilla) and "Changhaï" (Shanghai)were not found in the French space).

In sum, when the cosine values between place names are computed, higher cosine values indicating semantic similarity correlate with lower distances in miles indicating proximity[57]. Several factors that could attribute to these findings have been ruled out, including a bias towards certain place names (US cities or world cities), corpus (TASA or Encyclopedia) or language (English or French). This suggests that LSA can capture semantic information from which spatial information can be inferred.

8. Discussion

This paper started out showing the similarities between knowledge representation models developed in AI and how they relate to the models proposed in psychology. The focus in the paper was on one particular cognitively inspired language-based knowledge representation system, Latent Semantic Analysis. 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[47, 58] argued that this is due to the fact that LSA is a fundamentally amodal symbolic presentation. Others[46] would argue that this is due to the fact that LSA is deprived of perceptual information. Our results suggest that the claim that LSA cannot deal with "real world" simulations is debatable. LSA simulates categorization and prototypicality reasonably well. Furthermore, it ranks concepts of time in a temporally appropriate order. Finally, by just using higher-order co-occurrences LSA correlates with real distances between places, independent of the type of city (US cities or world cities), corpus (TASA or encyclopedia) or language (English or French). Important thereby is that a word is not compared to one other words, but that the item is compared in relation to all other items.

NL-KR systems have their 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. One problem not mentioned so far is the inability of LSA to conceptually combine the spatial or temporal structure of words together in discourse. That is, in LSA, the order of words in a sentence plays no part in the meaning that is derived from the words. For instance, *John kissed Mary*, is represented in the same way as *Mary kissed John* (cos = 1.0). This limitation currently poses a problem for LSA being a fully functioning model of language. Possible solutions to this problem lie in understanding how to combine syntactic features of particular words with their semantic content.

Despite these limitations, our results do raise doubts about the weaknesses of LSA's (and other NL tools) ability to capture spatial meaning between individual words. Space and time seem to be represented in LSA if a higher-order technique can make these temporal and spatial dimensions explicit, thereby providing new perspectives on, as well as new perspectives for, knowledge representations that are based in natural language.

Acknowledgments

This research was supported by NSF (IIS-0416128). 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.

9. References

1. M. Quillian. *Semantic memory* (Bolt, Beranek, & Newman, Cambridge, 1966).
2. A. Collins and M. Quillian, Retrieval time from semantic memory, *Journal of Verbal Learning and Verbal Behavior **8*** (1969), 240-248
3. E. Smith, E. Shoben, and L. Rips, Structure and process in semantic memory: A featural model for semantic decisions, *Psychological Review* **81** (1974) 214-241.
4. M. Minsky, A framework for representing knowledge. *The psychology of computer vision*, ed. P. Winston (New York: McGraw-Hill, 1975), 211-277.
5. D. Rumelhart and A. Ortony, The representation of knowledge in memory, *Schooling and the acquisition of knowledge*, eds. R. Anderson, R. Spiro and W. Montague, W. (Hillsdale: Lawrence Erlbaum, 1977), pp. 99-135.
6. R.C. Schank and R.P. Abelson, Scripts, plans, goals and understanding: an inquiry into human knowledge structures (Erlbaum, Hillsdale, 1977).
7. B.R. Gaines, Class library implementation of an open architecture knowledge support system, *International Journal of Human-Computer Studies* **41** (1994), 59-107.
8. D.G. Bobrow and T. Winograd, An overview of KRL, a knowledge representation language, *Cognitive Science* **1**, (1977), 3-46.
9. R.E. Fikes and T. Kehler, The role of frame-based representation in knowledge representation and reasoning, *Communications of the ACM* **28** (1985), 904-920.
10. R.J. Brachman, V. P. Gilbert and H. J. Levesque, An essential hybrid reasoning system, *Proceedings IJCAI-85* (1985), 532-539.
11. B. Nebel, Computational complexity of terminological reasoning in BACK, *Artificial Intelligence* **34** (1988), 371-383.
12. R.M. MacGregor, The evolving technology of classification-based knowledge representation systems, *Principles of semantic networks: Explorations in the representation of knowledge* . Ed. J.F. Sowa (Morgan-Kaufman, San Mateo, 1991), pp. 385-400.

13. A. Borgida, R.J. Brachman, D.L. McGuiness and L.A. Resnick, CLASSIC: a structural data model for objects, *Proceedings of 1989 SIGMOD Conference on the Management of Data* (ACM Press, New York, 1989), pp.58-67.
14. F. Baader and B. Hollunder, KRIS: Knowledge representation and inference system, *ACM SIGART Bulletin* **2** (1991), 8-14.
15. D.B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems* (Reading, Addison-Wesley, 1990).
16. F. Gomez (2000). Why base the knowledge representation language on natural language? *Journal of Intelligent Systems, 10.*
17. G.A. Miller, ed., WordNet: An on-line lexical database, *International Journal of Lexicography* **3** (1990), 235-312.
18. V. Rus and D.I. Moldovan, High precision logic form transformation, *International Journal on Artificial Intelligence Tools* **11** (2002), 437-454.
19. M. Pasca and S. Harabagiu, 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* (Carnegie Mellon University, 2001), 138-143.
20. A.C. Graesser, D.S. McNamara, M.M. Louwerse and Z. Cai, Coh-Metrix: Analysis of text on cohesion and language, *Behavior Research Methods, Instruments, and Computers* **36** (2004), 193-202.
21. W. Lewis, Measuring conceptual distance using WordNet: The design of a metric for measuring semantic similarity, R. Hayes, W. Lewis, E. Obryan, and T. Zamuner, eds., *The University of Arizona working papers in linguistics* (University of Arizona, 2002).
22. G. Vigliocco, D. Vinson, W. Lewis and M. Garrett, Representing the meaning of object and action words: The featural and unitary semantic space (FUSS) hypothesis, *Cognitive Psychology* **48** (2004), 422-488.
23. K. Lund and C. Burgess, Producing high-dimensional semantic spaces from lexical co-occurrence, *Behavior Research Methods, Instrumentation, and Computers* **28** (1998), 203-208.
24. T.K. Landauer and S.T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge, *Psychological Review* **104** (1997), 211-240.
25. X. Hu, Z. Cai, M.M. Louwerse, A. Olney, P. Penumatsa, A.C. Graesser and TRG Proceedings of the 2003 International Joint Conference on Artificial Intelligence (pp. 1489-1491) (2003).
26. X. Hu, Z. Cai, D. Franceschetti, P. Penumatsa, A.C. Graesser, M.M. Louwerse, D.S. McNamara, D.S., & TRG, LSA: The first dimension and dimensional weighting, *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, eds. R. Alterman and D. Hirsh (Cognitive Science Society, Boston, 2003), pp. 587-592.
27. M.W. Berry, S.T. Dumais, G.W. O' Brien, Using linear algebra for Intelligent Information Retrieval, SIAM Review **37** (1995), 573-595.
28. S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer and R. Harshman, Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*, 391-407 (1990).
29. S.T. Dumais in *The Second Text Retrieval Conference (TREC2)*, ed. Harman, D. (Natl. Inst. Stand. Technol., Gaithersburg, MD, 1994), pp. 105-116.
30. T.K. Landauer, P.W. Foltz and D. Laham, An introduction to latent semantic analysis. *Discourse Processes* **25** (1998), 259-284.
31. E. Kintsch, D. Steinhart, G. Stahl and the LSA Research Group, Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments* **8** (2000) 87-109.

32. A.C. Graesser, K. VanLehn, C. Rose, P. Jordan and D. Harter, Intelligent tutoring systems with coversational dialogue. *AI Magazine* **22** (2001), 39-51.
33. A.C. Graesser, P. Wiemer-Hastings, K. Wiemer-Hastings, D. Harter, N. Person and the TRG, Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments* **8** (2000), 128-148.
34. A.C. Graesser, S. Lu, G.T. Jackson, H. Mitchell, M. Ventura, A. Olney and M.M. Louwerse, AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers* **36** (2004), 180-193.
35. P.W. Foltz, W. Kintsch, and T.K. Landauer, The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, **25** (1998), 285-307.
36. W. Kintsch, Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review* **7** (2000), 257-266.
37. W. Kintsch (2002) On the notions of theme and topic in psychological process models of text comprehension, *Thematics : Interdisciplinary Studies*, M. Louwerse and W. van Peer, eds., (Benjamins, Amsterdam, 2002), pp. 157-170.
38. W. Kintsch, *Comprehension: A paradigm for cognition* (Cambridge University Press, New York, 1998).
39. M.M. Louwerse, Semantic variation in idiolect and sociolect: Corpus linguistic evidence from literary texts. *Computers and the Humanities* **38** (2004), 207-221.
40. M.M. Louwerse, P.M. McCarthy, D.S. McNamara and A.C. Graesser, Variations in language and cohesion across written and spoken registers, *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, K. Forbus, D. Gentner, T. Regier, eds. (Erlbaum, Mahwah, 2004), pp. 843-848.
41. D. Biber, Variations across speech and writing (Cambridge University Press, Cambridge, 1988).
42. J.F. Quesada, W. Kintsch and E. Gomez, Latent Problem Solving Analysis as an explanation of expertise effects in a complex, dynamic task, *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, eds. R. Alterman and D. Hirsh (Cognitive Science Society, Boston, 2003).
43. T.K. Landauer, Latent Semantic Analysis (LSA), a disembodied learning machine, acquires human word meaning vicariously from language alone. Commentary on Barsalou, *Behavior and Brain Sciences* **22** (1999), 624-625.
44. T.K. Landauer, On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The psychology of learning and motivation* **41** (2002), 43-84.
45. M.M. Louwerse and M. Ventura, How children learn the meaning of words and how computers do it (too). *Journal of the Learning Sciences* **14** (2005), 301-309.
46. L.W. Barsalou, Perceptual symbol systems. *Behavior and Brain Sciences* **22** (1999), 577-660.
47. A.M. Glenberg and D.A. Robertson, Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* **43** (2000), 379-401.
48. M.M. Louwerse, Iconicity in amodal symbolic representations, *LSA: A road to meaning*, T. Landauer, D. McNamara, S. Dennis & W. Kintsch, eds. (Erlbaum, Mahwah, in press).
49. T. Deacon, The symbolic species: The co-evolution of language and the human brain (Allen Lane, London, 1997)
50. C.S. Peirce, *The Collected Papers of Charles Sanders Peirce*, eds. C. Hartshorne, P. Weiss and A. Burks (Harvard University Press, Cambridge, 1931-58).
51. T. Deacon (1997: 90)
52. E. Rosch, On the internal structure of perceptual and semantic categories, *Cognitive Development and the Acquisition of Language*, T.E. Moore, eds. (Academic Press, New York, 1973), pp. 111-144

53. E. Rosch and C. Mervis, Family resemblances: Studies in the internal structures of categories, Cognitive Psychology **7** (1975), 573-605.
54. D. Laham, Latent Semantic Analysis approaches to categorization, *Proceedings of the 19th annual meeting of the Cognitive Science Society*, eds. M.G. Shafto and P. Langley (Mahwah, NJ: Erlbaum, Mahwah, 1997), p. 979).
55. M.M. Louwerse, Z. Cai, X. Hu, M. Ventura, P. Jeuniaux, The embodiment of amodal symbolic knowledge representations, *Proceedings of the 18th International Florida Artificial Intelligence Research Society*, eds. *I. Russell and Z. Markov* (AAAI Press, Menlo Park, 2005), pp. 542-47.
56. J.B. Kruskal and M. Wish, *Multidimensional scaling* (Sage University Series, Beverly Hills, 1978).
57. R.A. Zwaan, *Experiential traces and mental simulations in language comprehension* (under review).
58. M.M. Louwerse and P. Jeuniaux, *Language comprehension is both embodied* and *symbolic!* (under review)