Title
The Context Dependent Sentence Abstraction model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Ventura, Matthew
Hu, Xiangen
Graesser, Art
et al.

Publication Date
2004

Peer reviewed

The Context Dependent Sentence Abstraction model

Matthew Ventura (mventura@memphis.edu)
Department of Psychology/ Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152-3115 USA

Xiangen Hu (xhu@memphis.edu)
Department of Psychology/ Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152-3115 USA

Art Graesser (a-graesser@memphis.edu)
Department of Psychology/ Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152-3115 USA

Max Louwerse (mlouwers@memphis.edu)
Department of Psychology/ Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152-3115 USA

Andrew Olney (aolney@memphis.edu)
Department of Computer Science/ Institute for Intelligent Systems, 365 Innovation Drive
Memphis, TN 38152-3115 USA

Abstract

The Context Dependent Sentence Abstraction (CDSA) model and Latent Semantic Analysis (LSA) were compared in their ability to predict sentence similarity. Evidence supports the conclusion that the CDSA model better predicts human ratings for short phrases and sentences than does LSA. Alternative theoretical reasons are given for this finding.

Introduction

Researchers in many disciplines within cognitive science have proposed and tested theoretical claims about the meaning of natural language expressions. One of the contemporary models is Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). LSA is a statistical, corpus based technique for representing world knowledge. It computes similarity comparisons between words or documents by capitalizing on the fact that words are similar when they are surrounded by similar words (i.e., the company a word keeps).

LSA takes quantitative information about co-occurrences of words in documents (paragraphs and sentences) and translates this into a K-dimensional space. The input of LSA is a large co-occurrence matrix that specifies the frequency of words in documents. LSA reduces each document and word into a lower dimensional space by using singular value decomposition. This way, the initially extremely large word-by-document co-occurrence matrix is typically reduced to about 300 dimensions. Each word ends up being a K-dimensional vector. The semantic relationship between words can be estimated by taking the cosine (normalized dot product) between two vectors. Although LSA performance has been shown to be impressive at the paragraph level (Foltz, Gilliam, & Kendall, 2000; Landauer, Laham, Rehder, & Schreiner, 1997), other research has found limitations of LSA at the sentence level (Kintsch, 2001). In this paper we will present the Context Dependent Sentence Abstraction (CDSA) model, a corpus-based model that builds sentence meanings based on combinations of pooled adjacent neighbors of individual words. We will first discuss a weakness with vector representational systems (e.g., LSA) in handling sentence comprehension and then turn to a description of the CDSA model, with evidence supporting it.

A weakness with LSA

One major strength of LSA is its versatility and simplicity in handling word meaning and sentence meaning by the use of vector representations. It could be argued, however, that there are potential theoretical problems with combining word vectors to form sentences. For example, the meaning created from a sentence in LSA is a linear combination of word vectors, without eliminating information for any word. Consider the sentence *the cow ate in the field.* In LSA <u>all</u> information about *cows* (e.g., animal, milk, burger), *ate* (e.g., food, grocery, digest), and *field* (e.g., grass, baseball, football) may be included in the sentence representation. It could be argued that this assumption is not theoretically plausible because much of this associated information is not relevant to the word in context. There must be constraints that narrow down the vast array of information that may be "primed" in the first stages of sentence comprehension. Indeed, Kintsch's construction-integration model (1998) has attempted to explain this convergence of activated information by principles that guide the integration mechanisms.

Whereas the standard use of LSA is based on the assumption that a sentence's meaning is the sum of all the individual word meanings, there are extensions. Kintsch's predication algorithm (2001) tries to build meaning of a sentence by using syntactical information and LSA to create dependencies between subjects, predicates, and objects. For example, consider the sentences *the horse ran* and *the color ran*. The context established by *ran* has different meanings in these two sentences. Therefore, in the predication algorithm, constraints are made on what *ran* means in these sentences. The first step is to find the near neighbors of the word *ran* (i.e., words that give the highest cosine to *ran*). For the *horse* example, all the neighbors of *ran* are compared to the word *horse*. This provides words like *walk, gallop, crawl, rode,* etc. These neighbors of *ran* that are closest to *horse* (i.e., highest cosine) are then included into the vector for the sentence *the horse ran*. The same is done for the color example, resulting in different overall meanings. Including this additional information has been shown to more accurately capture the meaning of a sentence when we consider metaphor and causal inferences (Kintsch, 2001).

Kintsch's predication algorithm (2001) therefore imposes augmentations and constraints on the standard use of LSA. However, this algorithm still may not go the distance in solving the problem of information overload mentioned earlier. That is, predicating the verb *ate* to *cow* does give relevant information like *graze*, but all information about *cows* and *ate* are also included. To successfully implement context in the given example, we would want to include only information about "*cows eating"*, not about "*cows* and *ate* and *graze* and *field* and *pasture"*. While the predication algorithm solves some problems by adding information, it also may be limited by not taking any information away.

The need for contextual constraints

Computational representations like LSA go beyond general word meanings, but may not adequately handle contextual constraints. LSA may go some distance in handling proposition meanings that constrain words in context (Kintsch, 1998), but there still is a large landscape of representations and algorithms for combining information from words. We propose a new way of implementing contextual constraints. These contextual constraints are first built from simple individual word meanings that get established over time from their occurrences in the environment. But as sentences are constructed, similarities between the words in the constrained construction build a new meaning different from the sum of its parts.

The CDSA Model

Associationist frameworks (Landauer, 2002; Louwerse & Ventura, in press; Smith, Jones, & Landau, 1992) assume that it is critically important to measure and model the correlations between occurrences or events in the environment. We pursued a corpus-based model of word and sentence meaning, called the Context Dependent Sentence Abstraction (CDSA) model. In the CDSA model, semantic information within any word w is the pooled words that co-occur with word w in every context. One of the goals of this model is to try and capture the associations between words under a new level of specificity that considers the pool of their surrounding words.

In order to implement this model, it was necessary to make decisions about the learning rule and training set to be used. For this model, the deciding factor in each of these cases was psychological plausibility. That is, this model considers a corpus of prior experiences with words in context and the theoretical weights between words that change with experience, as opposed to a priori sets of features that are dictated by a brittle, symbolic model. The central question is how these weights change with experience. The proposed CDSA claims that they change by accumulating specific sentence exemplars.

Consider two words *chair* and *table*. The central question to be asked is what are all the possible <u>relevant</u> or <u>useful</u> relations that can exist between these two concepts? Each word has a neighborhood set that includes all words that co-occur with the target word. These words are the extensional meaning of the target word and serve as the basis for all associations. The neighborhood intersection is the relation that occurs when two words share similar co-occurrences with other words. Much like LSA, words become associated by their occurrence with many of the same words. For example, *food* and *eat* may become associated because they both occur with words such as *hungry* and *table*. Therefore the neighborhood set $\boldsymbol{N}$ for any word w is all the information we have in the exemplars for a word.

Neighbor weights

The neighborhood set for any word is intended to represent the meaning of a word from a corpus. But there were several theoretical challenges that arose when we developed the model. One dealt with how to differentially weight neighborhood words. We assigned *neighborhood weights* to each neighborhood word n of word w according to Equation (1).

$$\lambda_{n|w} = \frac{f(n \mid w)}{\sqrt{f(w)f(n)}} \quad (1)$$

The expression $f(n|w)$ designates the frequency of occurrence of the neighbor word n to target word w, whereas $f(n)$ is the total frequency of the neighbor word n, and $f(w)$ is the total frequency of the target word w. This formula essentially restricts the weights for the neighbor words as being between 0 and 1 in most cases. We adopted this simple assumption but we acknowledge that there are other ways to guarantee the range of the weights being within 0 and 1.

Therefore, the weighting function was aimed at giving more importance to words that consistently co-occur and less importance to words that occur frequently in the corpus. Additionally, rare co-occurrences may be given low weights because they do not consistently co-occur with the target word.

Some important assumptions had to be made in order to build relevant associations to target words most effectively. The next section will explain the procedures of the algorithm written to perform these operations.

Neighborhood Intersection Algorithm

In order to construct the neighborhood set for any word, an algorithm was written that pooled all words $\boldsymbol{N}$ that co-occurred with the target word w. We used the Touchstone Applied Science Associates (TASA) corpus because of its size (750,000 sentences) and diversity of topics (reading a diversity of texts up to college level). Each sentence in the corpus served as the context for direct co-occurrence. So for entire set of sentence sentences ($s_1 ... s_C$) that target word w occurs in, every unique word in ($s_1 ... s_C$) is pooled into the neighborhood set $\boldsymbol{N}$. For example the neighborhood of *chair* may consist of: *table, sit, leg, baby, kitchen, talk*, etc. This represents the neighborhood $\boldsymbol{N}$ of each target word w. Each word in the set ($n_1 ... n_K$) of $\boldsymbol{N}$ is weighted by the function described in equation (1). To evaluate the relation between any two words w_1 and w_2, we follow the following algorithmic procedure:

1. Pool neighborhood sets for w_1 and w_2 ($\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ respectively), computing the weights for all the neighbor words using Equation (1).
2. Calculate neighborhood intersection as follows:

$$\frac{\sum_{n \in N_1 \cap N_2} \left(\lambda_{n|w_1} + \lambda_{n|w_2} \right)}{\sum_{n \in N_1 \cup N_2} \left(\lambda_{n|w_1} + \lambda_{n|w_2} \right)} \qquad (2)$$

The numerator is the summation of weights over the intersection of the neighborhood sets ($\boldsymbol{N}_1$ and $\boldsymbol{N}_2$) whereas the denominator is the summation of weights over the union of the two neighborhood sets. This formula produces a value between 0 and 1.

In the next section we will discuss how the CDSA model was evaluated.

CDSA Model Evaluation

In four experiments we evaluated the CDSA model against LSA and human raters. The estimations of word and sentence meanings in the CDSA model and LSA were trained on the TASA corpus. Ratings in all four experiments were made by 10 undergraduate psychology students who were instructed to rate the similarity of various pairs of words (i.e., primarily from words from Spellman, Holyoak, & Morrison, 2001) on a 6-point scale that varied from 1 (very unrelated) to 6 (very related). A rating of 1 or 2 meant the rater could not easily find a functional or physical relationship between the word pairs (e.g. fish-office). The mean among the raters for each pair was taken as the basic data to test the models.

Experiment 1

Word Pairs A total of 64 word pairs was constructed that had a frequency over 10 in the TASA corpus. Some of the words were expected to be unrelated (e.g., chair-hear) and some related (e.g., chair-sit) in order to provide a sensitive range of values.

Results and Discussion

Human ratings (M = 3.57, SD = 2.20) were significantly correlated with the values produced by the CDSA model, $r = .71$, $p < .001$, and with LSA cosines, $r = .78$, $p < .001$. So both models fared quite well in accounting for the ratings of word pairs.

Neighborhood intersection estimation shared a relation to human ratings, so we might conclude that this type of association between words is used in human judgments. That is, by using all the co-occurrence information about a word, one can capture the meaning of a word. As can be seen, LSA was slightly more predictive of word relations than the CDSA model, although the difference was not statistically significant.

The lack of difference between models may be due to the construction of neighborhood sets for a single word in the CDSA model. Since there are many neighbors that exist for any particular word, there are many degrees of freedom that exist for determining the meaning for a single word. For instance, if one is asked to give an association to the word *cow*, there are many possible associations (e.g., *animal, milk, burger*, etc), which will lead to a very general non-specific representation of a single word.

The purpose of Experiment 2 is to try to use the model to represent the meaning of word-pairs. This involves imposing constraints on the neighbors for each pair in order to more accurately represent the contextual meaning of the pair. For instance, *cow-graze* should give a more specific representation of *cow* than *cow* without a context because constraints are built on the meaning of *cow*. These constraints initially involve measuring the neighborhood overlap between the neighbors of *cow* and the neighbors of *graze*, which then are used to compare to another set of information (e.g., word, sentence).

Experiment 2

A central theoretical assumption in Experiment 1 was the idea that neighborhood intersection plays a prominent role in the relation between words. But how can the current model account for conceptual relationships beyond the word level? Figure 1 gives an illustration of how this could be done. If two pairs are being compared, the neighborhood overlap of each pair is pooled into $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$. Then the intersection (Equation 2) is calculated to access the similarity between the two pairs. This constrains the degrees of freedom for the pair, which eliminates any information that is not mutually shared by both words in the pair (i.e., the problem found in Kintsch's predication algorithm). Therefore, each word is always dependent on the context in which it appears. As the context for a word becomes more specific (i.e., as reflected by the number of unique words it appears with), the less likely that the same context will be associated with any random word. For instance, *chair-sit* has a smaller neighborhood set than the sum of neighbors for *chair* and the neighbors for *sit*. This assumption therefore states that word pairs, or even sentences, are different than the sum of its parts, an assumption quite different from current models of associative learning like LSA.

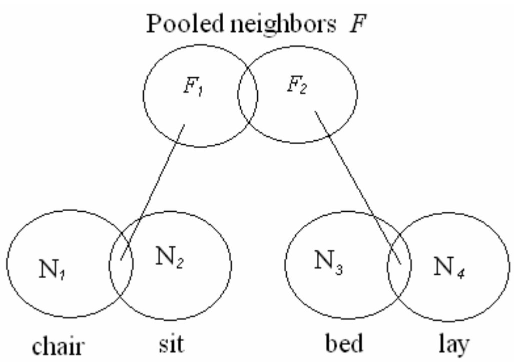

Figure 1: The recursive nature of neighborhood overlap. The neighborhood overlap ($\boldsymbol{F_1}$) of *chair* ($\boldsymbol{N_1}$) and *sit* ($\boldsymbol{N_2}$) is intersected with to the neighborhood overlap ($\boldsymbol{F_2}$) of *bed* ($\boldsymbol{N_3}$) and *lay* ($\boldsymbol{N_4}$).

Model modifications

Neighborhoods were first built on words within the pair as described in equation 1. As described in Figure 1, the neighborhood $\boldsymbol{N_1}$ of *chair* is intersected with the neighborhood $\boldsymbol{N_2}$ of *sit* to yield a new neighborhood $\boldsymbol{F_1}$ that represents the "chair sit" neighborhood. Since any shared neighbor in $\boldsymbol{N_1}$ and $\boldsymbol{N_2}$ each have a separate weight, the average of the two weights (Equation (3)) is calculated to represent the new weight for each neighbor in $\boldsymbol{F_1}$.

$$\lambda_{n|w_1 w_2} = \frac{1}{2}\left[\lambda_{n|w_1} + \lambda_{n|w_2}\right] \qquad (3)$$

In the same manner, we obtain $\mathbf{F_2}$. Once $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$ have been calculated for both word-pairs, the neighborhood intersection is calculated (as described in Equation (2)), to access the relationship between the 2-word pairs.

Additionally the union of the neighborhood weights (i.e., the entire neighbor weights of all words in each pair) was calculated for $\boldsymbol{F}$ to compare the effectiveness of intersecting the neighborhoods.

Word Pairs We constructed 53 word pairs that had a frequency over 10 in the TASA corpus. Separate sets of pairs were intended to be unrelated (e.g., bear/cave—pen/write), related by analogy (e.g., bear/cave—fish/pond), or related by both analogy and semantic relation (e.g., teeth/bite—leg/kick).

Results and Discussion

Human ratings (M = 3.46, SD = 1.62) were significantly correlated with CDSA intersection, r = .60, p < .001, and union, r = .51, p < .001. LSA cosines were also related to human similarity ratings, r = .64, p < .001.

It appears that imposing context reduced the correlation with rated similarity of 2-word pairs, compared with single-word pairs. As can be seen LSA performance also drops. Most notably, the union of neighbor sets does not perform as well as the intersection version of the CDSA model.

The purpose of Experiment 3 was to examine how performance would be affected by implementing more context through comparison of 3-word phrases.

Experiment 3

The process we used in building constraints on three-word combinations involves a multinomial neighborhood overlap (N-O) among all neighborhood pairs. Each neighbor that is shared by at least two neighborhoods is then pooled into $\boldsymbol{F}$. Figure 2 gives an illustration of how this can be achieved..

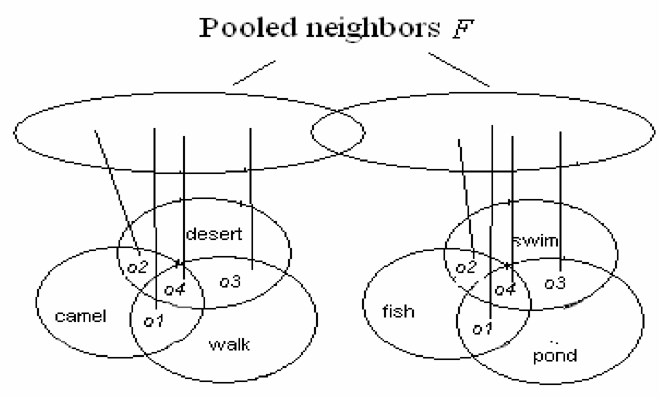

Figure 2: 3-word neighborhood overlap for two 3-word combinations. The neighbors of each combination are compared for neighborhood overlap, which are then pooled into a neighborhood $\boldsymbol{F}$ for each pair.

Model Modifications

Equation 3 is used to compute weights for the words that are in the intersection of any two neighborhoods (o_1, o_2, and o_3).

By making all possible intersections between each neighborhood $\boldsymbol{N_1}$, $\boldsymbol{N_2}$, and $\boldsymbol{N_3}$, a select number of words may be counted three times. Therefore neighbors shared by all three sets (o_4; see figure 2) will be averaged (i.e., divided by 3) and eliminated from any other N-O to avoid multiple counts. The computation for the weights in the intersection of the three neighborhoods is simply an extension of equation 3, where the average is taken with three weights instead of two. Additionally each neighbor in o_4 is a special neighbor because it is shared by all neighbor sets. Therefore these neighbors are multiplied by a constant of 3 (i.e., since there are three sets) to give greater importance to these context bound neighbors. Once $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$ have been pooled together by all the N-O for each pair, the neighborhood intersection is calculated (Equation 2), to access the relationship between the 3-word pairs.

Additionally the union of the neighborhood weights (i.e., the entire neighbor weights of all words in each pair) was calculated for $\boldsymbol{F}$ to compare the effectiveness of intersecting the neighborhoods.

Short phrases We constructed 58 three-word phrases that had a frequency over 10 in the TASA corpus. Some pairs were intended to be unrelated (e.g., bird/nest/fly—brush/paint/art), related by analogy-like relations (e.g., gun/shot/bullet — axe/chop/wood), and related by both analogy and semantic relation (e.g., dog/loud/bark—cat/quiet/meow).

Results and Discussion

Human ratings (M = 3.10, SD = 1.78) were significantly related to the CDSA intersection, $r = .63$, $p < .001$, and union, $r = .44$, $p < .001$. LSA cosines were related to human similarity ratings, $r = .47$, $p < .001$. The results give evidence that imposing context improves performance in calculating similarity. Furthermore, LSA performance continues to drop as more word context is introduced. This in part could be due to the lack of constraints that are put in the sentence representation in LSA.

Experiment 4

The purpose of the present experiment is to test the CDSA model to sentences of varying lengths (i.e., sentences ranging from 4 to 6 words). One challenge that arises in calculating sentence similarities is how to handle all the possible intersections between word neighbors within one sentence. Therefore three conditions were tested on how to calculate the final sentence neighbor set $\boldsymbol{F}$. First, a multinomial approach entailing N-O among all neighborhood sets was pooled to get $\boldsymbol{F}$. Weightings were computed between any N-O words as an extension of Equation 3, where the neighborhood intersections could entail 2-6 neighborhoods.

Second, a word-chunking maximum likelihood approach was used that calculated a set $\boldsymbol{P}$ for every three words in a sentence (Johansson, 2000). This chunking approach using a 3-word context to any target word was found to give equal performance to 5-word and 7-word contexts in syntactic tagging. So if a sentence had five words, a multinomial N-O calculation between the 1st 2nd and 3rd word neighborhoods would produce $\boldsymbol{P_1}$ (i.e., as described in Experiment 3), then N-O would be calculated between the 4th and 5th neighborhood words to produce $\boldsymbol{P_2}$ (as described in Experiment 2). Then the N-O between $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ would be calculated to produce the final neighborhood $\boldsymbol{F}$ for the sentence. The intersection between $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$ (Equation 2) will give the final similarity between the two sentences. This word-chunking hypothesis is consistent with the intuition that adjacent words in a sentence constrain meaning more than nonadjacent words in a sentence.

Finally, the union of the neighborhood weights (i.e., the entire neighbor weights of all words in each pair) was calculated for $\boldsymbol{F}$ to compare the effectiveness of intersecting the neighborhoods.

Sentences We constructed 42 sentences whose words had a frequency over 10 in the TASA corpus. Sentences pairs were constructed of varying length (e.g., *blue bird fed babies nest tree -- bear protected cubs den*; articles, pronouns and prepositions were removed). Sentences were constructed so that about half were considered related and half unrelated.

Results and Discussion

Human ratings (M = 2.12, SD = 1.48) were significantly correlated to CDSA model 3 word chunking intersection, $r = .69$, $p < .001$, CDSA union, $r = .65$, $p < .001$, and the CDSA multinomial intersection, $r = .56$, $p < .001$. LSA cosines were also related to human similarity ratings, $r = .50$, $p < .001$.

The results give evidence that imposing context may be important when calculating sentence similarity. By applying an arbitrary rule set to sentences of varying lengths seems to yield better performance than just making all possible intersections among neighbors. Alternatively, the union of all the neighbors seems to perform just as well as a rule based intersection procedure. Possible reasons for this will be discussed next.

General Discussion

In sentence comprehension, comprehenders must understand the nature of word context and the constraints one word places on another (Kintsch, 1998). In other words, comprehenders will have to ask themselves: how does the meaning of one word affect the meaning of another word? The most straightforward relationship among words is an additive one, where the meaning of one word has no influence on the meaning of another word. In contrast, in the case of sentence comprehension, the levels of a one word can dramatically change the effects of another word. In this model, context refers to N-O among words in a sentence. That is, changing levels of one word can dramatically affect the meaning of another word. Thus, without structural constraints involving processes similar to

N-O, sentence meanings proceed in a radically different manner. Many relations shared between the pairs in the 4 experiments were abstract relations, ones that were only clearly established by filtering the individual word meanings and keeping shared information among words.

The word-chunking N-O approach appears to perform better than the multinomial N-O approach among all neighbors. Making all possible intersections among neighbors does not seem to be very psychologically plausible since it would involve making many comparisons between words that may not be relevant. For instance, comparing the first word to the last word in a sentence may not be important in evaluating the meaning of a sentence.

Possible improvements

As can be seen in experiment 4, N-O did not seem to help predict sentence similarity to a great extent over the union of all the neighborhoods in a sentence. This may be due to the arbitrariness of the rules used to calculate N-O for varying sentence lengths. For instance, if the sentence was 6 words long, N-O would be calculated for the three words and the last three words. With these two pools we would then calculate $\boldsymbol{F}$. This type of rule makes the assumption that all 6-word sentences follow the same syntactic structure. This obviously will not do for all 6-word sentences. Therefore, it seems likely that if the CDSA model was implemented with a syntactic parsing mechanism, it could give the correct word pairs to calculate N-O for any sentence.

Conclusion

The computational model presented here captures both word and sentence meaning. There are several reasons why using the CDSA model is advantageous. First, it uses simple mechanisms that are psychologically plausible. Second, it gives the freedom to add more information to the corpus at any time. Since the measures derived are computed on-line on the corpus, dynamically adding text to the corpus is not a problem. Essentially, many weights are changed between words as soon as text is added.

The proposed computational model captures word and sentence meaning by appealing to constraints reflected in a corpus analysis. Embodiment theorists (Glenberg & Robertson, 2000) may claim that there is no meaning derived from a corpus analysis because the words are not grounded in sensory-motor experience. In principle, one could have a more grounded corpus with units extensively embedded in sensory and motor experience. The TASA corpus was simply readily available. Whether the episodic experiences are reflected in TASA or in sensory-motor experience, the theoretical assumptions of the CDSA model are that, specific exemplars and associative processes are sufficient to account for the judgments of meaning similarity. The CDSA model uses simple mechanisms that rely on co-occurrences of words in exemplars.

One additional advantage of the CDSA model is that it allows more information to be added to the corpus at any time. Since the measures derived are computed on-line on the corpus, dynamically adding text to the corpus is not a problem. Essentially, weights are changed between words as soon as text is added.

Acknowledgments

This research was supported by the National Science Foundation (REC 0106965, ITR 0325428) and the DoD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ONR or NSF.

The TASA corpus used was generously provided by Touchtone Applied Science Associates, Newburg, NY, who developed it for data on which to base their Educators Word Frequency Guide.

References

Foltz, P.W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in on-line writing evaluation with LSA. *Interactive Learning Environments, 8*, 111-127.

Johansson, C. (2000). A Context Sensitive Maximum Likelihood Approach to Chunking. In: *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.

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

Kintsch, W. (2001) Predication. *Cognitive Science* 25, 173-202.

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., Laham, D., Rehder, B., & Schreiner, M. E., (1997). How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual meeting of the Cognitive Science Society* (pp. 412-417). Mahwah, NJ: Erlbaum.

Landauer, T. K. (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.

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*.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, 60, 143-171.

Spellman, B. A., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory & Cognition, 29*, 383-393.