Semantic and Associative Priming in High-Dimensional Semantic Space

Kevin Lund
Department of Psychology
University of California, Riverside
Riverside, CA 92521
kevin@locutus.ucr.edu

Curt Burgess
Department of Psychology
University of California, Riverside
Riverside, CA 92521
curt@cassandra.ucr.edu

Ruth Ann Atchley
Department of Psychology
University of California, Riverside
Riverside, CA 92521
ruthann@moriarty.ucr.edu

Abstract

We present a model of semantic memory that utilizes a high-dimensional semantic space constructed from a co-occurrence matrix. This matrix was formed by analyzing a 160 million word corpus. Word vectors were then obtained by extracting rows and columns of this matrix. These vectors were subjected to multidimensional scaling. Words were found to cluster semantically, suggesting that interword distance may be interpretable as a measure of semantic similarity. In attempting to replicate with our simulation the semantic and associative priming experiment by Shelton and Martin (1992), we found that semantic similarity plays a larger role in priming than what they would suggest. Vectors were formed for three different types of related words that may more orthogonally control for association and similarity, and interpair distances were computed for both related and unrelated prime-target pairs. A priming effect was found for pairs that were only semantically related, as well as for word pairs that were both semantically and associatively related. No priming was found for word pairs which were strictly associatively related (no semantic overlap). This finding was replicated in a single-word priming experiment using a lexical decision procedure with human subjects. The lack of associative priming is discussed in relation to prior experiments that have found robust associative priming. We conclude that our priming results are driven by semantic overlap rather than by associativity, and that prior results finding associative priming are due, at least in part, to semantic overlap within the associated word pairs.

Introduction

In cognitive psychology the concept of the semantic network was introduced by Collins and Quillian (1969). In their model, semantic meaning or concepts were represented by nodes which corresponded to individual words. Links connected these nodes as a function of the type of relationship they shared, and each link varied in length reflecting the strength of the relationship. An outcome of such a model is that processing one concept will facilitate the subsequent processing of a related concept. Meyer and Schvaneveldt (1971) provided an early demonstration of how recognizing semantically related words (BREAD-BUTTER) can speed the lexical decision latencies compared to seeing unrelated words (FLOOR-BUTTER). Semantic memory research utilizing the lexical decision (or naming) task in the last two decades has produced one of the largest bodies of cognitive psychological literature (see Neely, 1991). The presence of the semantic priming effect is one of the most robust effects in the literature, although the exact nature of the word relationships and the nature of the task and methodology can influence the magnitude and presence of priming (Fischler, 1977; Neely, 1991).

In general, however, regardless of the theory, information processed at a semantic level facilitates related semantic information (semantic features: Smith, Shoben, & Rips, 1974; nodal networks: Collins & Quillian, 1969; prototype models: Rosch, 1973; associative relationships: Lupker, 1984; distributed connectionist representations: McClelland & Kawamoto, 1986; or mental models: Johnson-Laird, 1983). Attempting to derive models of semantic memory that would allow the computation of a distance between two concepts and that could be used to reflect semantic relatedness using psychometric techniques has a long history in cognitive psychology dating back at least to Osgood, et al. (1957). A common approach has been to use multidimensional scaling on (many thousands of) human judgments of similarity (Smith, Shoben, & Rips, 1974; Schvaneveldt, 1990). More recently, investigators using large scale corpora have attempted to extract semantic information directly from text. Gallant (1991) has developed a methodology that extracts a distributed set of semantic microfeatures utilizing the context in which a word is found. However, a drawback to his approach is that the features for the core meaning have to be determined by a human judge.

The model we have developed accounts for a wide range of semantic effects in the cognitive and neuropsychological literature (Burgess & Lund, in press; Lund & Burgess, in press). In this paper we focus on the facilitation effect, known as semantic priming, that is found when a word is preceded by a semantically related word. We refer to our model as HAL, for Hyperspace Analogue to Language. We use a large text corpus of 160 million words to initially track lexical co-occurrence within a 10 word moving window. From the co-occurrences, we develop a cognitively plausible 200 dimensional semantic space by using the most informational dimensions. The development of our model was strongly influenced by the work of Schütze (1992) who has developed a similar model for information retrieval. Our goal has been to incorporate cognitive constraints into a high-dimensional semantic space model and to extend this

Cognitive Science Proceedings, LEA. (1995)

approach to determine if it would account for the semantic phenomena found in the cognitive science literature.

Four experiments are presented here. First we demonstrate the semantic nature of the word vectors that we extract using HAL. We then use the model to generate word vectors for three types of related word stimuli from two previously published studies and show that the semantic relatedness effect is present for the two categories of words that share semantic features and is absent for the words that are associatively, but not semantically, related. In the final experiment, with human subjects, we replicate the effect we found with the model.

Simulation Methods

Matrix construction

The basic methodology of the simulation is to develop a matrix of word co-occurrence values for a given vocabulary. This matrix will then be divided into co-occurrence vectors for each word, which can be analyzed for semantic content.

An analysis of co-occurrence must define a window size; that is, the largest number of words that may occur between a pair of words such that the pair may be considered to co-occur. The limiting case of a small (useful) window is a width of one, which would correspond to counting only immediately adjacent words as co-occurrants. At the other end of the spectrum, one may count all words within a logical division of the input text as co-occurring equally (see Landauer, 1994; Schvaneveldt, 1990). A very small window may miss constructs spanning several words (lengthy noun phrases, for instance), while large windows risk introducing large numbers of extraneous co-occurrences. Therefore, we chose a window width of ten words. Our hopes are that this preserves locality of reference, while obscuring the effects of different syntactic constructions. This syntax independence may be important when comparing results from different languages. As a further move away from dependence on syntax (or any structuring of the language under consideration other than that given by the division of words), sentence boundaries are ignored.

Within this ten-word window, co-occurrence values are inversely proportional to the number of words separating a specific pair. A word pair separated by a nine-word gap, for instance, would gain a co-occurrence strength of one, while the same pair appearing adjacent to one another would receive an increment of ten. Cognitive plausibility was a constraint, and a ten-word window with decreasing co-occurrence strength seemed within these bounds (Gernsbacher, 1990).

The product of this procedure is an N-by-N matrix, where N is the number of words in the vocabulary being considered. It is this matrix which we will demonstrate to contain significant amounts of semantic information.

Text source. The corpus that was analyzed was approximately 160,000,000 words of English text gathered from Usenet. All newsgroups containing English dialog were included. This source has a number of properties which we found appealing. First, it is voluminous. It was clear that in order to obtain reliable data across a large vocabulary, a large amount of text would be required. Usenet was attractive in that it could supply several million words of text per day, indefinitely. Second, Usenet is diverse. Virtually no subject goes undiscussed, which allows the construction of a very broadly based co-occurrence data set. This turns out to be useful when attempting to apply the data to various stimulus sets. There is little chance of running across words or word senses which were not encountered during matrix construction. Third, the text is conversational. Rather than the formal business reports or specialized dictionaries found in other corpora, Usenet text resembles everyday speech more closely than most corpora. That the model works with noisy, conversational input suggests that it can robustly deal with some of the same problems that the human-language comprehender encounters.

Vocabulary. The vocabulary used for the analysis consisted of the 70,000 most frequently occurring symbols within the corpus. A check against the standard Unix dictionary showed that only one half of these were valid English words. The abundance of nonword symbols is not a practical problem, though, since the data-reduction step involves extracting information and frequent nonword symbols (including slang words and misspellings) presumably carry useful information.

Data reduction. The co-occurrence tabulation produces a 70,000 by 70,000 matrix. Each row of this vector represents the degree to which each word in the vocabulary preceded the word corresponding to the row, while each column represents the co-occurrence values for words following the word corresponding to the column. A full co-occurrence vector for a word consists of both the row and the column for that word. The following experiments operate on groups of these co-occurrence vectors. To reduce the amount of data involved, the column variances of the particular vectors used in each experiment are computed, and the columns with he smallest variances are discarded. We find that variance drops sharply across the first hundred elements, and is very low by the two hundredth element; accordingly, the 139,800 columns with the lowest variance are discarded. This leaves a 200-element vector for each word. Empirically, these shortened vectors provide similar results to the full-length vectors, while being much easier to work with.

These vectors (whether length 140,000 or 200) can be viewed as the coordinates of points in a high-dimensional space, with each word occupying one point. Using this representation, differences between two words' co-occurrence vectors can be measured as the distance between the high-dimensional points defined by their vectors.

Experiment 1

Methods

A number of words, informally chosen to represent three categories (animal types, body parts, and geographical locations) were used in an initial analysis of semantic content for the co-occurrence vectors. A vector of length 200 was extracted for each word, and, treating each vector as a set of coordinates in a 200-dimensional Euclidean space, a distance matrix was formed. Our hypothesis was that this distance matrix, representing the inter-point distances for the chosen set of words, would operate as a similarity matrix. Each element in the matrix represented the distance between two of the chosen words in the high-dimensional space.

This matrix was analyzed by a multidimensional scaling algorithm (MDS), which projects points from a high-dimensional space into a lower-dimensional space, in a non-linear fashion that attempts to preserve the distances between points as much as possible. The lower-dimensional projection allows us to visualize the spatial relationships between the co-occurrence vectors. The two-dimensional MDS solution is shown in Figure 1.

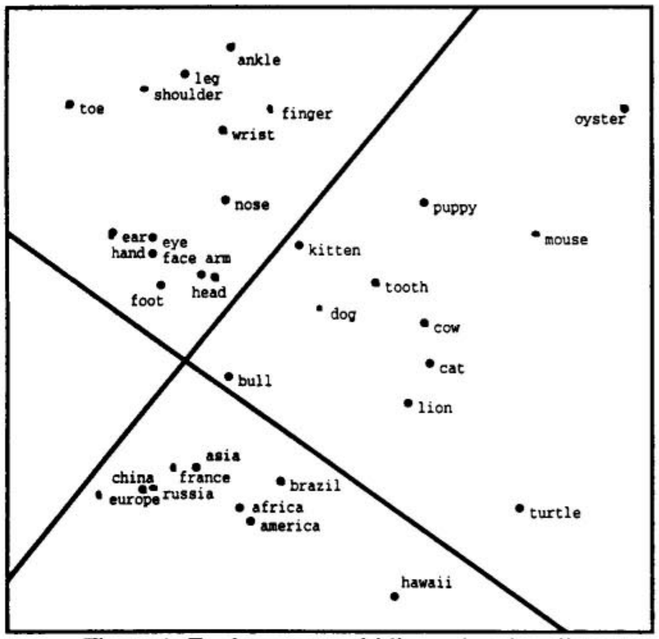

Figure 1: Exploratory multidimensional scaling

Results & Discussion

Visual inspection suggests that the three groups of words were differentiated by this procedure. Geographic regions are distinct from the other two groups, while animals and body parts overlap at "tooth" (a salient body part for animals).

Given that words with similar meanings tended to be close to each other in the simulation data, we conclude that the most informational vector elements from the co-occurrence matrix carry semantic information. However, recent research suggests that so-called semantic priming may, in fact, be carried by associative relationships rather than semantic similarity (Shelton & Martin, 1992). In Experiment two our goal was to simulate a pattern of semantic and associative priming results found in the literature using the semantic vectors extracted with HAL and to determine the correspondence of the vector information to semantic and associative characteristics of words.

Experiment 2

Subject responses to words that are preceded by words related to them are facilitated, presumably by the prime having lowered the recognition threshold for the related target. Shelton and Martin (1992) found that automatic priming using a lexical decision task does not occur for items that are only semantically related, yet does occur when items are both associatively and semantically related. This pattern of results held for only the single presentation procedure in which a subject makes a lexical decision to each word as they are presented one after another, rather than the standard single-word priming procedure where there is a discrete trial usually consisting of a fixation point, a prime word, followed by a target. Shelton and Martin attribute semantic priming to controlled processes, such as subject expectancy. Thus, they conclude, automatic priming will obtain only for items which are associatively related and not for items that are only semantically related.

The results from Experiment 1 suggest that the distances between words are semantic in nature and that the distances might roughly correspond to reaction time latencies from experiments like Shelton and Martin's. In this experiment we attempt to replicate Shelton and Martin's Experiment 4, using their stimuli, in which they find priming for the associated items, but not for the semantically related items. We reason that since we are calculating priming effects using HAL's vector representations that we have eliminated the possibility of attentional processing and strategic effects that Shelton and Martin suggest are necessary for true semantic priming.

Methods

The relatedness or priming effect for an item was calculated by subtracting the distance of the related pair from an unrelated word pair (using the same target). Semantic vectors of length 200 were formed, and for each word pair the Euclidean distance between the words was computed. Results are shown in Table 1. Distances are given in Riverside Semantic Units (RSUs), which are are completely arbitrary and based upon normalized vectors.

Table 1: Semantic Distances for Shelton & Martin

	Related	Unrelated	Priming
Semantic	366	429	63
Associated	310	407	97

Results & Discussion

Table 1 presents the semantic distances and priming effects. In the analysis of variance, there was a main effect for relatedness, demonstrating that the semantic distances for all related pairs was shorter than that for the unrelated pairs, $F(1,70) = 15.43$, $p = .0001$. There was not an interaction, $F<1$. These results are not completely consistent with those found by Shelton and Martin who did not find a priming effect for the semantically related items. We suggest several possibilities for this inconsistency. First, a methodological reason. The single presentation procedure in all their experiments resulted in overall faster reaction times than the single-word priming procedures. A number of investigators have demonstrated that semantic effects tend to be minimized or disappear with either faster responses or faster subjects (Hines, et al., 1986; Chiarello, et al., 1990).

A more theoretical concern has to due with the stimuli. Both the semantically and the associatively related items used by Shelton and Martin share semantic characteristics, but more importantly, according to Shelton and Martin, is that the word pairs in the semantic condition are not associated. A closer look at their stimuli (see Table 2 for examples and semantic distances) suggests that the two conditions may not be semantic related to the same degree. We suggest that the semantic condition had a number of weak examples of related items (e.g., PEAS-GRAPES, MAID-WIFE), whereas, the associated condition had a number of items which were strongly semantically related (e.g., ROAD-STREET, GIRL-BOY). The means of semantic distances from Table 1, further suggest this is the case (semantic related, 366; associated related, 310) with the associated condition being marginally more semantically similar, $F(1,70) = 3.63$; $p = .061$. Experiment 3 is an attempt to more orthogonally manipulate the stimulus constraints of semantic relatedness and associativity.

Table 2: Stimuli from Shelton & Martin Experiment

	Semantic			Associated		
	Word pair		Dist.	Word pair		Dist.
Good	nose	hand	340	coffee	cup	575
Pairs	car	wagon	348	light	lamp	542
Weak	maid	wife	580	road	street	170
Pairs	peas	grapes	541	girl	boy	215

Experiment 3

Chiarello, Burgess, Richards, and Pollack (1990) used three types of word pairs to try and differentiate the contribution of semantic similarity and association in priming. Examples of the three types of words (semantically similar words, associated words, and words which are both semantically similar and associatively related) can be seen in Table 3. Semantically related words (TABLE-BED) are instances of the same category and share a number of features. Associated words (MOLD-BREAD) are those which are associated as determined by human word-association norms and tend to co-occur in sentential phrases. These items, however, are not instances of the same category and therefore share few semantic features. The third type of word relation are pairs that are both semantically and associatively related (UNCLE-AUNT). These word relations should allow us to distinguish between the associative and semantic components of our "semantic" vectors.

Semantically similar pairs were of the same superordinate semantic category and, thus, are likely to occur in similar contexts. Word co-occurrence or simple association reflects the temporal nature of language, whereas semantics pertains more to the internal features of words. Spence and Owens (1990) found that the production likelihood in word-association norms correlated with word co-occurrence. Thus, if the vectors generated by HAL are semantic in nature, we should see shorter distances between the word pairs of both semantic conditions than between word pairs in the associated condition. Alternatively, the nature of the semantic relationships could more directly correspond to the co-occurrence association. In such a case, we would expect a similar magnitude of prime-target distance in all three relatedness conditions.

Table 3: Example word pairs (Experiments 3 and 4)

Semantic		Associated		Both	
table	bed	cradle	baby	ale	beer
music	art	mug	beer	uncle	aunt
flea	ant	mold	bread	ball	bat
circle	cross	waist	belt	sofa	chair
pan	bowl	nest	bird	butter	bread

Methods

This experiment used the word pairs from Chiarello et al. (1990), which are divided into the three relatedness groups. The relatedness or priming effect for an item was calculated by subtracting the distance of the related pair from an unrelated word pair (using the same target).

Results and Discussion

The left side of Table 4 presents the semantic distances, priming effects, and standardized priming effects. In the analysis of variance, there were two factors: type of relation (semantic, associated, semantic and associated) and relatedness (related, unrelated). There was a main effect for relatedness, demonstrating that the semantic distances for related pairs was shorter than that for the unrelated pairs, $F1,270) = 11.62$, $p = .0008$. There was also a main effect for type, $F(2,270) = 4.24$, $p = .015$. There was not an interaction, $F(2,270) = 1.25$, $p = .28$. Planned comparisons were made at each level of word relation in order to determine priming by stimulus type. Visual inspection of Table 4 suggests that the priming effect for the associated trials was much less than the effect for both of the semantically related conditions. The analysis of variance is consistent with this observation. Priming is found for the semantic condition, $F(1,88) = 6.48$, $p = .012$, and the semantic and associated condition, $F(1,92) = 5.79$, $p = .018$, but not for the associated only condition, $F < 1$.

Standardized scores are computed by dividing by the semantic-only score. This pattern of results is consistent with the hypothesis that the vectors associated with each word do carry semantic information, but do not as strongly carry the associative information that would reflect the temporal characteristics inherent in words that co-occur. Our claim, to be discussed in more detail later, is that the word vectors are semantic in nature, even though their origin is from a co-occurrence matrix that tracked temporal sequence.

It would be important to see if this pattern of results can

Table 4: simulation distances (in RSUs and ms)

	HAL simulation			Human subjects		
	Sem.	Assoc.	Both	Sem.	Assoc.	Both
R	347	322	331	643	623	603
U	413	339	391	673	634	631
U-R	66	17	60	30	11	28
Std.	1.0	0.26	0.91	1.0	0.36	0.93

be replicated with human subjects. Chiarello et al. (1990), whose stimuli we utilized, used a divided visual-field methodology since they were primarily interested in priming in the cerebral hemispheres. In Experiment 4, we use these stimuli in a single-word priming study with human subjects, centrally presenting all stimuli.

Experiment 4

Methods

Sixty-four undergraduate students participated in order to earn course credit. The stimuli used in this experiment included the items used in Experiment 3, as well as, an equal number of word-nonword trials, since a lexical decision task was employed. The target words were balanced for both word length and printed frequency for a total of 288 word pairs. An experimental list included these trials and was preceded by four "warm up" trials. Word primes were counterbalanced so that a target would be preceded by a related word in one list and an unrelated word in a second list. This allowed the targets to act as their own controls. Of the related word-word trials, one third were word pairs that were only semantically related, one third were only associatively related, and a third were semantically and associatively related.

Stimulus presentation and timing was conducted on 486 PCs. Each trial began with a 500 ms fixation cross, followed by a prime at this location for 300 ms immediately followed by the target which remained until either the subject made a lexical decision or 2500 ms elapsed. Accuracy feedback was provided, along with a time-out signal for responses over 2500 ms.

Results and Discussion

The reaction-time latencies, priming effects, and standardized priming effects are presented in the right half of Table 4. Error rates were under 3% and showed no interesting variation. This analysis parallels the analysis presented in Experiment 3. Reaction times were faster to related trials than to unrelated trials, $F(1,270) = 14.70$, $p= .0002$. There was also a main effect for type, $F(2,270)=16.14$, $p=.0001$. There was not an interaction, $F < 1$. Planned comparisons were made at each level of word relation in order to determine priming by stimulus type. Inspection of Table 4 suggests that the pattern of reaction-time priming is similar to that found in Experiment 3, i.e., the priming for the associated trials was less than that for both the trials in the semantically related conditions. Priming obtains for the semantic condition, $F(1,88) = 5.82$, $p = .017$, and the semantic and associated condition, $F(1,92)= 12.38$, $p=.0007$, but not for the associated only condition, $F(1,90) = 1.14$, $p = .28$.

The results of Experiment 4 are virtually identical to those of Experiment 3. The similar pattern of results suggests that the semantic vectors that are extracted from the corpus are cognitively plausible, and that they incorporate higher-level semantic information that may, in part, correspond to semantic category and semantic feature similarity.

General Discussion

Several important conclusions can be drawn from this series of experiments. The basic relatedness or priming effect so often demonstrated in the memory literature can be reproduced using the semantic vectors from the HAL simulation. This priming effect, however, was not a general effect that was seen with just any set of related word pairs. A crucial aspect to the relatedness effect was that it seems to hinge more on the semantic relationships between words than on the associative relationships. There was no reliable effect of associativity. The nature of our associative items in Experiments 3 and 4 was that the targets were strong associates of the primes. Intuitively, there is a strong contiguous aspect to these stimuli. For example, MOLD-BREAD and CRADLE-BABY are word pairs that seem likely to occur together in a phrase or sentence. Historically in the literature on associativity, temporal order or contiguity has been conceded to be the most important principle in learning (Deese, 1965). We concede that first-order association is important in learning, however, our view is that these first-order temporal associations are not an important part of structural semantics. We want to make clear, though, that we consider second-order association, that is, the patterns of intercorrelations amongst word use, to be a principal building block of semantic structure. It is something akin to these intercorrelations that we propose that we extract with the 200 most variant, or informational, elements in our semantic vectors for each word.

We have demonstrated that in order to show a priming effect in the simulation, semantic similarity is required. An examination of the word pairs suggests at least a partial explanation for this. Semantically similar word pairs are interchangeable within a sentence; the resulting sentences may be pragmatically improbable, but they are not nonsensical (sentences 1a and 1b). Associated-only pairs tend to produce awkward sentences when interchanged, sentences that often cannot be taken literally (sentences 2a and 2b). The semantic vectors take us beyond simple co-occurrence in that they are really measures of context. Being interchangeable, the semantically similar words tend to appear in similar contexts, and so have similar vector representations.

1a) The child slept on the **bed**.
1b) The child slept on the **table**.
2a) The child slept in the **cradle**.
2b) The child slept in the **baby**.

Being "associated," without semantic similarity, was not sufficient in a word pair to produce priming in the simulation. The same pattern of results obtained with the human subjects. Obtaining reliable associated priming is sometimes difficult (see Chiarello, et al., 1990; Shelton & Martin, 1992, Exp. 3).

The vector-based semantic distances, being ultimately derived from human linguistic output, can be construed as human-based estimates of word similarity. Fischler (1977) found a correlation of 0.31 between human estimates of semantic similarity and facilitation in a priming experiment.

This finding closely parallels our correlation of 0.24 between semantic distances and reaction time. Thus, we can conclude that HAL's semantic distances are reliable predictors of lexical-decision latencies. Fischler further found a correlation of essentially zero (-0.01) between his assessment of associative relation and facilitation, which matches our finding of little to no priming in the associated-only case (with both simulation and humans).

There is something of a paradox when comparing the present results with a number of findings in the literature. For example, although Fischler (1977) found no correlation between the magnitude of priming and an index of associativity with his associated stimuli, he did, unlike us, find a priming effect with the trials. We suspect that the priming effect was a function of the semantic relationships of his word pairs. When we extracted semantic vectors for his associated stimuli we found that the three pairs of words with the closest semantic vector representations carried 1/3 of the priming effect. These items such as ROAD-STREET, are similar to our semantic and associated items. Similarly, Shelton and Martin (1992) found that associated items resulted in priming, whereas semantically similar, but non-associated, items did not. Again, though, as they note, their "associated" items possess considerable semantic similarity (e.g., GIRL-BOY, QUEEN-KING; see Table 2), not unlike the situation in Fischler's experiment. We would expect a robust priming effect with stimuli like these, but not as a result of the associative component. Shelton and Martin's semantically related items (e.g., DIRT-CEMENT, SOUP-JUICE), we suspect, are simply not similar enough to produce priming in a bottom-up fashion.

A major goal of this line of research was to develop a model of semantic memory from the analysis of a large language corpus. A fundamental methodology for investigating semantic organization is the priming experiment. One of the first obligations of any such model must be to account for the range of priming effects found in the literature. HAL has succeeded in producing a set of results that make a clear distinction between two types of information inherent in word relationships. We conclude that semantic information underlies the word priming effect and that first-order associative information is not as crucial as previously thought.

Further, the results suggest a theoretical explanation for the difficulty in finding associated-only priming. If a major organizing principle of semantic memory is the analysis of immediate word context, words will be organized by semantic similarity rather than associativity, as in our simulations. So far, our evidence shows this to be the case.

A final conclusion is that we have presented a methodology that exploits the regularities of language in a large text corpus such that the extraction of semantic information is possible. The methodology employed in HAL does not require supervised learning or other system feedback and works on very noisy, speech-like, input. A limitation to previous models of semantic processing is that either the semantic representations required extremely time-consuming human judgments about the items or the semantic representations were simply conjectural. HAL is a model that provides the generation of semantic representations of real words used in real language.

Acknowledgments

This research was supported by a NSF Presidential Faculty Fellow award SBR-9453406 and a U.C. Academic Senate Research Grant to Curt Burgess. We thank Chad Audet and Dr. Catherine Decker for their assistance.

References

Burgess, C., & Lund, K. (in press). Modeling the cerebral asymmetries of semantic memory in high-dimensional space: HAL's brain (abstract). *Brain and Cognition*.

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't . . .sometimes, some places. *Brain & Language*, 38, 75-104.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore: John Hopkins Press.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, 5, 335-339.

Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: LEA.

Hines, D., Czerwinski, M., Sawyer, P.K., & Dwyer, M. (1986). Automatic semantic priming: Effect of category exemplar level and word association level. *Journal of Experimental Psychology*: Human Perception and Performance, 12, 370-379.

Landauer, T.K., & Dumais, S. (1994). *Memory model reads encyclopedia, passes vocabulary test*. Paper presented at the Psychonomics Society.

Lund, K., & Burgess, C. (in press). Hyperspace Analogue to Language (HAL): A general model of semantic representation (abstract). *Brain and Cognition*.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-235.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner and G.W. Humphrey (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: LEA.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana: University of Illinois Press.

Schütze, Hinrich (1992). Dimensions of Meaning. *Proceedings of Supercomputing*.

Schvaneveldt, R.W. (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Shelton, J.R., & Martin, R.C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1191-1210.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.

Spence, D.P., & Owens, K.C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19, 317-330.