

## The Dynamics of Meaning in Memory

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

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

"Semantics. The curse of man."  
Maxwell (1976, p. 19)

"... how a word 'stands for' a thing or 'means'  
what the speaker intends to say or 'communicates'  
some condition of a thing to a listener has never  
been satisfactorily established"  
B. F. Skinner (1957, pp. 114-115)

"... semantic structure of natural languages  
evidently offers many mysteries"  
Noam Chomsky (1965, p. 163)

Meaning provides the fundamental bridge between the various language, cognitive, and perceptual components of the language comprehension system. As such, it is important to attempt to model how meaning can be acquired from experience and the specific nature of its representational form. In this chapter, we attempt to deal with the particularly difficult problem of how meaning can be specified. In particular, we are interested in how meaning can be represented in a computational model of meaning and the process by which these representations are formed. Although a review of previous models of word meaning would be outside the scope of this chapter (but see Komatsu, 1992), there are three psychological models that deserve mention because they have inspired current

---

This research was supported by a NSF Presidential Faculty Fellow award SBR-9453406 to Curt Burgess. Catherine Decker, Art Markman, Sonja Lyubomirsky, and X anonymous reviewers provided many helpful comments and we want to thank Jeff Elman for providing his corpus. More information about research at the Computational Cognition Lab, a HAL demo, and reprint information can be found at <http://HAL.ucr.edu>. Correspondence should be addressed to Curt Burgess, Psychology Department, 1419 Life Science Bldg., University of California, Riverside, CA 92521-0426. E-mail: [curt@cassandra.ucr.edu](mailto:curt@cassandra.ucr.edu).

computational approaches in many ways. Collins and Quillian (1969; 1972; also Collins & Loftus, 1975) developed a hierarchical network model of semantic memory. It is a node and link model where knowledge is represented by both concepts (the nodes) and the relations among concepts (the links). Superordinate and subordinate relationships (hence, the hierarchical nature of the model) are represented via the links. The later version of the model, the spreading activation model (Collins & Loftus, 1975), de-emphasized the hierarchical nature of the mental representations in favor of a more general notion of semantic relatedness. The information retrieval process occurs as a function of spreading activation in the structured network. There has been considerable support for the model; the spreading activation approach to meaning retrieval and representation has been extensively used (see Neely, 1991, for a review). The notions of semantic connectedness, spreading activation, perceptual thresholds for conceptual retrieval are present in many more contemporary localist connectionist models (Burgess & Lund, 1994; Cottrell, 1988).

Smith, Shoben, and Rips (1974), in their feature comparison model, hypothesized that there were two types of semantic features: defining features that were essential to the meaning of the concept and characteristic features that were usually true of the concept. Processing in this model hinged on whether an overall feature comparison or only a comparison using defining features was required for a semantic decision. The processing characteristics of both the spreading activation model and the feature comparison model have been better described than has their representational characteristics.

A very different approach to developing a semantic system was taken by Osgood and his colleagues (Osgood, 1941; 1952; 1971; Osgood, Suci, & Tannenbaum, 1957). Their work is likely the most ambitious attempt to empirically derive a set of semantic features. Osgood pioneered the use of the semantic differential in developing a set of semantic indices for words. With this procedure, a person rates a word using a likert scale against a set of bipolar adjective pairs (e.g., *wet-dry*, *rough-*

*smooth, angular-rounded, active-passive*). For example, the concept *eager* may get rated high on *active* and intermediate on *wet-dry*. The meaning of a word, then, would be represented by this semantic profile of ratings on a set of adjectives. The aspect of meaning represented by each adjective pair is a dimension in a high-dimensional semantic space. Distances between words in such a space essentially constitutes a similarity metric that can be used to make comparisons among words or sets of words. An advantage of the semantic differential procedure is that all words have coordinates on the same semantic dimensions making comparisons straightforward. A drawback to the procedure is that it requires considerable overhead on the part of human judges. In one study reported by Osgood, et al. (1957) 100 likert scale judgments were collected for each of the 50 adjective scales for 20 words. Thus, 100,000 human judgments were required for a set of semantic features for these 20 words. Human semantic judgments were used by Rips, Shoben and Smith (1973) who had people make typicality judgments on a small set of words in order to generate a two-dimensional semantic representation. Although meaning-based models can be developed by using judgments about word meaning, the effort is extensive for even a small set of words. Both the semantic differential and word association norms (Deese, 1965) share the problem that there is considerable human overhead in acquiring the information. There is, perhaps, a more serious problem at a theoretical level. As Berwick (1989) has argued, selecting semantic primitives is a "hazardous game" (p. 95). These different procedures do not begin to deal with issues such as how word meaning acquisition occurs, the role of simple associations in learning more general knowledge, a mechanism for linking environmental input to the form of a mental representation, the relationship between episodic and semantic representations, and the creation of abstract representations.

### **Representing Meaning in Computational Models**

The use of semantic representations in computation models very much corresponds to that found with the psychological models just discussed. In this section, three means of representing semantic features will be described that encompass most computational approaches.

The spreading activation model of Collins and Loftus (1975) and the feature comparison model of Smith, et al. (1974) provide the inspiration for many aspects of contemporary connectionist models. Feature vectors representing meaning can be found in

distributed connectionist models (Hinton & Shallice, 1991; McClelland and Kawamoto, 1986; Plaut & Shallice, 1994). In these models, the semantic features are specifically delineated (humanness, shape, volume, etc). The limitation of these connectionist models, however, is that there is usually just an intuitive rationale for the semantic features. For example, McClelland and Kawamoto used a set of distributed representations in their model of thematic role assignment and sentence processing. Words were represented by a set of semantic microfeatures. Nouns, for instance, had features such as *human*, *softness*, *gender*, and *form*. Verbs had more complex features such as *cause* (whether the verb is causal), or *touch* (specifies whether the agent or instrument touches the patient). This model was important in that it demonstrated that distributed semantic representations can account for case-role assignment and handle lexical ambiguity. Similar approaches to feature designation have been frequently used in the connectionist literature for more basic models of word recognition (Dyer, 1990; Hinton & Shallice, 1991; Plaut & Shallice, 1994).

A more empirically derived set of semantic features was developed by McRae, de Sa, and Seidenberg (1997). McRae, et al. had 300 subjects list what they thought were features to 190 words. This procedure resulted in a total of 54,685 responses. In their experiments they found that these feature representations and the pattern of inter-correlations among them predicted the pattern of behavioral priming results for natural kind and artifact categories. These feature lists were also used as the source for word vectors in a connectionist model of word representation.

Masson (1995) used a different approach in his model of semantic priming. Rather than have vector elements correspond to any actual aspect of meaning, he simply used 80 element word vectors such that related words had more elements that matched than unrelated words. Thus, his semantic vectors only indicated a degree of similarity between two items, not any particular relationship since the vectors are inherently "non-meaningful." The vector representations make no commitment to a particular set of features or theory of meaning, although the vector representations imply a certain degree of relatedness in order to model cognitive effects.

All three approaches use binary vectors; in some cases the vector elements correspond to specific featural aspects of word meaning, in other cases, it is simply the proportion of similar elements that dictate the general relatedness of word meaning. All these approaches have certain advantages in developing

models of meaning in that they are straightforward to set up and work well in complex learning models. What is not clear, however, is what features one would select for a more general model of semantic representation (beyond some small set of items) or for concepts that are abstract in nature. A drawback to developing a set of features from human feature list norms is that many human responses are required for each word of interest. This is not unlike the semantic differential technique in which the experimenter must choose the semantic dimensions upon which words are rated and then gather a large number of human judgments. Still, these approaches do seem to facilitate the development of processing models.

Gallant (1991) has attempted to extract semantic information directly from text using large-scale corpora. He 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 meanings have to be determined by a human judge.

The limitation of all these approaches (although less so with the feature list procedure) is that the nature of the representations does not foster much evolution of representational theory. Given the theoretical and computational importance of developing some principled set of meaning features it is surprising that so little has been attempted in deriving such a set. The Hyperspace Analogue to Memory (HAL) model to be discussed next does not rely on any explicit human judgments in determining the dimensions that are used to represent a word (other than deciding that the word is the unit) and acquires the representations in an unsupervised fashion. The model learns its representations of meaning from a large corpus of text. The concept acquisition process, referred to as global co-occurrence, is a theory of how simple associations in context are aggregated into conceptual representations. Memory is not a static collection of information -- it is a dynamic system sensitive to context. This dynamic relationship between environment and representation provides the basis for a system that can essentially organize itself without recourse to some internal agent or "self." The HAL model is a model of representation. As presented in this chapter HAL models the development of meaning representations, and, as implemented here, is not a process model.<sup>1</sup> The primary goal of this

<sup>1</sup> There are some exceptions to the statement that HAL is not a processing model which we will discuss later. We have implemented HAL as a processing model of cerebral asymmetries (Burgess & Lund,

Table 1. Sample Global Co-occurrence Matrix for the Sentence "the horse raced past the barn fell."

	barn	horse	past	raced	the
barn		2	4	3	6
fell	5	1	3	2	4
horse					5
past		4		5	3
raced		5			4
the		3	5	4	2

*Note:* The values in the matrix rows represent co-occurrence values for words which preceded the word (row label). Columns represent co-occurrence values for words following the word (column label). Cells containing zeroes were left empty in this table. This example uses a five-word co-occurrence window.

chapter is to address a series of critical issues that a dynamic model of memory must confront when providing a representational theory. We will argue that the HAL model provides a vehicle that has caused us to rethink many of the assumptions underlying the nature of meaning representation.

### The HAL Model

*Words are slippery customers.*

Labov (1972)

Developing a plausible methodology for representing the meaning of a word is central to any serious model of memory or language comprehension. We use a large text corpus of ~300 million words to initially track lexical co-occurrence within a 10-word moving window. From the co-occurrences, we develop a 140,000 dimensional context space (see Lund & Burgess, 1996, for full implementational details). This high-dimensional context or memory space is the word co-occurrence

(1998) and as a model of concept acquisition (Burgess, Lund, & Kromsky, 1997). Chad Audet is developing one of the lab's newest initiatives: a connectionist model that includes HAL context vectors for a meaning component along with phonology and orthography.

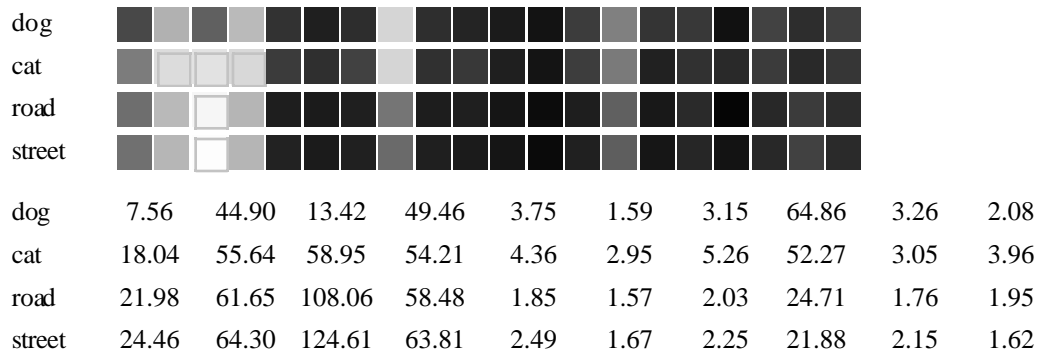


Figure 1. Sample 20-element word vectors for four words. Each vector element has a continuous value (the normalized value from its matrix cell) and is gray-scaled to represent the normalized value with black corresponding to zero. Below the gray-scaled vectors are the normalized numeric representations for the first 10 vector elements.

matrix. We refer to this high-dimensional space as a "context" space since each vector element represents a symbol (usually a word) in the input stream of the text. Each symbol is part of the textual context in the moving window.

*Constructing the Memory Matrix.* The basic methodology for the simulations reported here is to develop a matrix of word co-occurrence values for the lexical items in the corpus. This matrix will then be divided into co-occurrence vectors for each word, which can be subjected to analysis for meaningful content. For any analysis of co-occurrence, one must define a window size. The smallest useable window would be a width of one, corresponding to only immediately adjacent words. 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 & Dumais, 1994, 1997; Schvaneveldt, 1990).

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 adjacently would receive an increment of ten. Cognitive plausibility was a constraint, and a ten-word window with decreasing co-occurrence strength seemed a reasonable way to mimic the span of what might be captured in working memory (Gensbacher, 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 contains significant amounts of information that can be used to simulate a variety of cognitive phenomena. A sample matrix is

shown in Table 1. This sample matrix models the status of a matrix using only a 5-word moving window for just one sentence, *the horse raced past the barn fell*. An example may facilitate understanding this process. Consider the word *barn*. The word *barn* is the last word of the sentence and is preceded by the word *the* twice. The row for *barn* encodes preceding information that co-occurs with *barn*. The occurrence of the word *the* just prior to the word *barn* gets a co-occurrence weight of 5 since there are no intervening items. The first occurrence of *the* in the sentence gets a co-occurrence weight of 1 since there are 4 intervening words. Adding the 5 and the 1 results in a value of 6 recorded in that cell. This example uses a five word moving window; it is important to remember that the actual model uses a 10 word window that moves through the 300 million word corpus.

*Characteristics of Corpus.* The corpus that serves as input for the HAL model is approximately 300 million words of English text gathered from Usenet. All newsgroups (~3,000) containing English text were included. This source has a number of appealing properties. 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 indefinitely supply about twenty million words of text per day. In addition, Usenet is conversationally diverse. Virtually no subject goes undiscussed; this allows the construction of a broadly based co-occurrence data set. This turns out to be useful when attempting to apply the data to various stimulus sets since there is little chance of encountering a word not in the model's vocabulary. One goal for HAL was that it would develop its representations from conversational text that was minimally preprocessed,

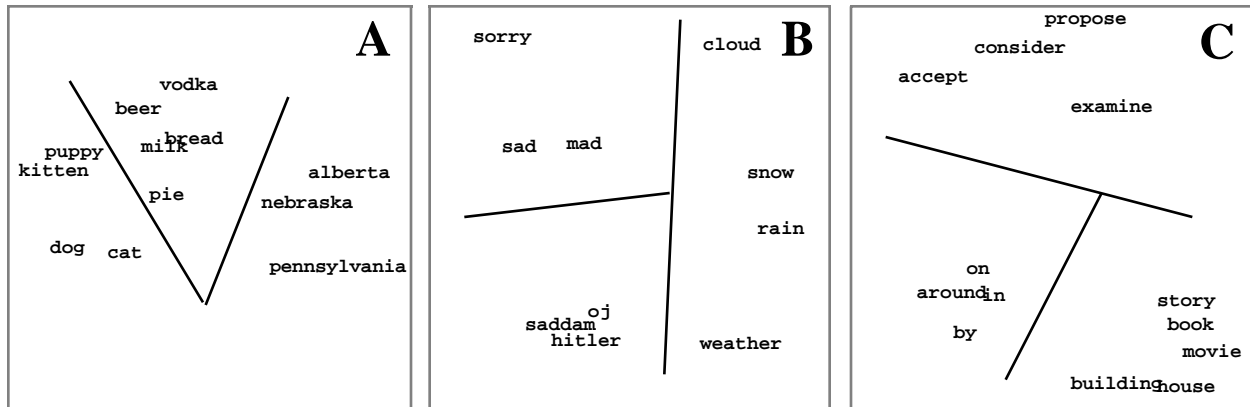


Figure 2. Two-dimensional multidimensional scaling solutions for : (A) common nouns, (B) abstract words, and (C) grammatical categories.

not unlike human-concept acquisition. Unlike formal business reports or specialized dictionaries that are frequently used as corpora, Usenet text resembles everyday speech. That the model works with such noisy, conversational input suggests that it can robustly deal with some of the same problems that the human-language comprehender encounters.

*Vocabulary.* The vocabulary of the HAL model consisted of the 70,000 most frequently occurring symbols within the corpus. About half of these had entries in the standard Unix dictionary; the remaining items included proper names, slang words, nonword symbols and misspellings. These items also presumably carry useful information for concept acquisition

*Data extraction.* 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 use groups of these co-occurrence vectors. These vectors (length 140,000) 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 (distance is measured in Riverside Context Units, or RCUs, see Lund & Burgess, 1996).

*Vector Properties.* As described above, each element of a vector represents a coordinate in high-dimensional space for a word or concept, and a distance metric applied to these vectors presumably

corresponds to context similarity (not just item similarity; this will be discussed more later). The vectors can also be viewed graphically as can be seen in Figure 1. Sample words (e.g., *dog*, *cat*) are shown with their accompanying 20 element vectors (only 20 of the 140,000 elements are shown for viewing ease). Each vector element has a continuous numeric value (the frequency normalized value from its matrix cell). A grey-scale is used to represent the normalized value with black corresponding to a zero or minimal value. The word vectors are very sparse; a large proportion of a word's vector elements are zero or close to zero. A word's vector can be seen as a distributed representation (Hinton, McClelland, & Rumelhart, 1986). Each word is represented by a pattern of values distributed over many elements, and any particular vector element can participate in the representation of any word. The representations gracefully degrade as elements are removed; for example, there is only a small difference in performance between a vector with 140,000 elements and one with 1,000 elements. Finally, it can be seen that words representing similar concepts have similar vectors, although this can be subtle at times (see Figure 1). See Lund and Burgess (1996) for a full description of the HAL methodology.

The HAL model has been used to investigate a wide range of cognitive phenomena. The goal of this chapter is to address a series of issues that are central to any theory of memory representation, rather than discuss any particular cognitive phenomenon in detail. As a precursor to that, Figure 2 was prepared to illustrate a variety of categorization effects that the HAL model has been used to investigate. In later sections, the primary literature where the more extensive results can be found will be referred to, but in the interim, Figure 2 can serve as a conceptual

starting point. The results in Figure 2 are analyses of stimuli from earlier papers using a multidimensional scaling algorithm (MDS) which projects points from a high-dimensional space into a lower-dimensional space in a non-linear fashion<sup>2</sup>. The MDS attempts to preserve the distances between points as much as possible. The lower-dimensional projection allows for the visualization of the spatial relationships between the global co-occurrence vectors for the items. Figure 2a is an example of how the vector representations carry basic semantic information that provides for the categorization of animals, foods, and geographic locations. Within category semantics can be seen as well. Alcoholic liquids cluster together in the food group; young domestic animals cluster separately from the more common labels (*dog*, *cat*). Distances between items have been used to model a variety of semantic priming experiments (which will be discussed in the next section). The stimuli in Figure 2b illustrate a particular feature of HAL's meaning vectors, namely, that they can be used to model abstract concepts that have been notably problematic for representational theory. Abstract concepts such as weather terms, proper names and emotional terms all segregate into their own meaning spaces. One advantage of representing meaning with vectors such as these is that, since each vector element is a symbol in the input stream (typically another word); all words have as their "features" other words. This translates into the ability to have a vector representation for abstract concepts as easily as one can have a representation for more basic concepts (Burgess & Lund, 1997b). This is important, if not absolutely crucial, when developing a memory model that purports to be general in nature. The other major aspect of categorization that the HAL model can address is the grammatical nature of word meaning. A clear categorization of nouns, prepositions, and

---

<sup>2</sup> Visual inspection of the MDS presentations in this paper all appear to show a robust separation of the various word groups. However, it is important to determine if these categorizations are clearly distinguished in the high-dimensional space. Our approach to this is to use an analysis of variance that compares the intragroup distances to the intergroup distances. This is accomplished by calculating all combinations of item-pair distances within a group and comparing them to all combinations of item-pair distances in the other groups. In all MDS presentations shown in this paper, these analyses were computed, and all differences discussed were reliable.

verbs can be seen in Figure 2c. The generalizability of the HAL model to capture grammatical meaning as well as more traditional semantic characteristics of words is an important feature of the model (Burgess, 1998; Burgess & Lund, 1997a) and was part of our motivation to refer to the high-dimensional space as a context space rather than a semantic space.

These and other characteristics of word meaning that the model encodes has led us to rethink a number of assumptions about the dynamics of memory and concept acquisition which will be addressed in the following sections. The HAL model offers a clearly defined way to think about what an association is in the learning process and the relationship of basic associations to higher-order word meaning. The grammatical characteristics encoded in the word vectors provokes a reconsideration of syntactic constraints and representational modularity. The global co-occurrence mechanism at the heart of the model provides the vehicle for rethinking what is meant by similarity. We think that HAL offers a more general statement about similarity than other models. One result of how the global co-occurrence mechanism works has allowed a proposal of how high-dimensional memory models can address the failure of previous computational model to deal with the symbol grounding problem. The role of context is central to all these issues that we will address. In one section, a comparison is made of the HAL implementation of a context based model and a recurrent neural network implementation. The similarity of the results of these two very different implementations makes a strong case for the strength of the contextual constraint in language input in forming conceptual representations. We now turn to the evidence for these arguments.

### **Rethinking the Nature of Associations**

In the HAL model, an association and semantic or categorical knowledge is clearly defined. These operational definitions can be used to shed light on an ongoing controversy in the priming literature as to what is meant by "semantic" priming and under what conditions it is obtained. Critical to this discussion is a distinction between semantic and associative relationships. In most experiments, word association norms are used to derive stimuli. However, word norms confound semantic and associative relationships. *Cat* and *dog* are related both categorically (they are similar animals) and associatively (one will tend to produce the other in production norms). The typical assumption behind associative relationships is that associations are caused by temporal co-occurrence in language (or

Table 2. Example prime-target word pairs from the Semantic, Associated, and the Semantic+Associated relatedness conditions.

Semantic		Associated		Semantic + Associated	
table	bed	cradle	baby	ale	beer
music	art	mug	beer	uncle	aunt
flea	ant	mold	bread	ball	bat

*Note:* The full set of these stimuli was taken from Chiarello, Burgess, Richards, & Pollock (1990).

elsewhere in the environment). Stimuli can be constructed such that these semantic-categorical and associative relationships can be, for the most part, orthogonally manipulated. To illustrate, *cat* and *dog* are semantically and associatively related. However, *music* and *art* are semantically related, but *art* does not show up as an associate to *music* in word norms. Conversely, *bread* tends to be one of the first words produced in norms to the word *mold*. However, *bread* and *mold* are not similar - clearly, though, this is not to say there is no relationship between *bread* and *mold*; they are just very different items. As the story goes, *mold* and *bread* would be likely to co-occur. Examples of these types of word pairs can be seen in Table 2.

*Semantic and Associative Priming.* Our claim is that HAL encodes experience such that it learns concepts more categorically. Associative - more episodic - relationships will have been aggregated into the conceptual representation. This can be seen by re-examining Table 1. The vector representation for *barn* will include the row and column of weighted co-occurrence values for the words that co-occurred with *barn* in the moving window. The representation for *barn*, as it stands in Table 1, is episodic. *Barn* has occurred in only this one context. As more language is experienced by HAL, the vector representation for *barn* accrues more contextual experience; and, as a result, the weighted co-occurrences sum this experience, resulting in a more generalized representation for *barn*. This is an important aspect of HAL for attempting to model priming. It follows that the distances in the hyperspace should be sensitive to more generalized, categorical relationships. Furthermore, the more associative relationships should not have a strong correlation to HAL's distance metric. We tested these hypotheses in two experiments (Lund, Burgess, & Atchley, 1995) using the three different types of word relationships illustrated in Table 2. These word relationships have various combinations of semantic and associative properties -- semantic only, associative only, and combined semantic and associative properties. There

is considerable research that shows that human subjects are sensitive to all three of these types of word relationships (Lund, et al. 1995; Lund, Burgess, & Audet, 1996; see Neely, 1991). We replicated that finding - subjects made faster lexical decisions to related word trials (in all three conditions) than to the targets in the unrelated pairs (Lund, et al. 1995). In a second experiment, we computed the context distance between the related and unrelated trials in all three conditions using HAL. Priming would be computed in this experiment by using the distances; there should be shorter distances for the related pairs than for the unrelated pairs in the representational model. In this experiment we found robust priming for the semantic-only and the semantic-plus-associative conditions. There was no distance priming in the model for the associated-only pairs. This result raises some intriguing questions about the representational nature of words and the ongoing controversy in the priming literature as to what is meant by "semantic" priming and under what conditions it is obtained.

The controversy exists, in part, due to a mixed set of results in the literature - some investigators obtaining semantic priming without association, others not finding semantic-only priming in conditions that would seem to limit strategic processing. Fischler (1977) has one of the earliest findings showing that strength of association did not correlate with priming. Similarly, Chiarello, Burgess, Richards, and Pollock (1990) found semantic-only priming using a low proportion of related trials and a naming task. However, Lupker (1984) did not find priming for semantically related word pairs that were not also associatively related. A similar set of results is found in Shelton and Martin (1992). They used a single presentation lexical decision task where words were presented one after another with lexical decisions made to each word. Such a procedure masks the obviousness of prime - target relations to a subject. Shelton and Martin did not find semantic priming under these conditions. A comparison of experiments such as these usually entails a comparison of the methodologies. Experiments that do not obtain

semantic-only priming typically avoid the lexical decision task, unless it is part of the individual presentation procedure (i.e., Shelton & Martin). The naming task is thought to be less sensitive to strategic effects (although this may also limit its sensitivity to semantic relations as well). Clearly experimental procedures and task differences play a part in these results. Focusing on task differences, however, may divert attention from important representational issues that are likely just as important. In developing representational theory, it is important not to make representational conclusions based solely on procedural issues.

We have argued that an experiment's sensitivity in reflecting the semantic-only priming effect is guided by the strength of the semantic (contextual) relationship (Lund, et al. 1995; 1996). One set of stimuli that we have evaluated in detail using the HAL model are the items used by Shelton and Martin (1992). We found that many of their semantic pairs (e.g., *maid-wife*, *peas-grapes*) were not closely related by using HAL's semantic distance metric. Furthermore, a number of their semantic and associated pairs were very strongly related categorically (e.g., *road-street*, *girl-boy*) (see Lund, et al. 1995). Using HAL, we argued that the semantic-only condition did not produce priming simply because the prime-target pairs in that condition were not sufficiently similar.

There are two experiments that offer compelling evidence that increased similarity results in priming under task constraints usually associated with a lack of semantic-only priming. Cushman, Burgess, and Maxfield (1993) found priming with the semantic-only word pairs used originally by Chiarello, et al. (1990) with patients who had visual neglect as a result of brain damage. What is compelling about this result is that the priming occurred when primes were presented to the impaired visual field. These patients were not aware that a prime had even been presented, thus making it difficult to argue for any strategic effect. A more recent result by McRae and Boisvert (1998) confirmed our earlier hypothesis generated by our HAL simulation that Shelton and Martin's (1992) failure to find priming was due to insufficient relatedness in their semantic-only condition. Recall that they used an individual-presentation lexical decision methodology. McRae and Boisvert replicated this methodology but used a set of non-associatively related word pairs that subjects rated as more similar than Shelton and Martin's items. McRae and Boisvert replicated Shelton and Martin with their items, but, using the more similar items, found a robust semantic-only priming effect. Thus, it appears that

increased attention to the representational nature of the stimuli affords a more complete understanding of the semantic constraints as well as the methodological issues involved in priming.

HAL's distance metric offers a way to evaluate stimuli in a clearly operationalized manner. The ~70,000 item lexicon provides the basis for which the stimuli from various experiments can be evaluated directly. In most experiments, word association norms are used to derive stimuli, and it is important to realize that word norms confound semantic and associative relationships.

We argue that HAL offers a good account of the initial bottom-up activation of categorical information in memory. It provides a good index of what information can be activated automatically. Although others have argued that it is associative, not semantic, information that facilitates the automatic, bottom-up activation of information (Lupker, 1984; Shelton & Martin, 1992), some of the confusion is a result of the field not having a clear operational definition of what an association is and how "an association" participates in learning. On one hand, an association is operationally defined as the type of word relationships that are produced when a person free associates. Yet this is an unsatisfying definition at a theoretical level since it divorces the acquisition process from the nature of the representation. It also confounds many types of word relationships that can be found using a word-association procedure.

*Word Association Norms.* One intuitive conception of word association is that it is related to the degree to which words tend to co-occur in language (Miller, 1969). Spence and Owens (1990) confirmed this long-held belief empirically. To see if this relationship between word association ranking and lexical co-occurrence held for the language corpus that we use for HAL, we used 389 highly associated pairs from the Palermo and Jenkins (1964) norms as the basis for this experiment (Lund, et al. 1996). We replicated Spence and Owens' effect; word association ranking was correlated (+.25) with frequency of co-occurrence (in the moving window). Our correlation was not as strong as theirs probably due to the fact that we used only the five strongest associates to the cue word. However, using all strongly associated word pairs allowed us to test a further question. To what extent is similarity, at least as operationalized in the HAL model, related to this co-occurrence in language for these highly associated words? We divided these strongly associated pairs into those that were semantic neighbors (associates that occurred within a radius of 50 words in the hyperspace) and those that were non-neighbors (pairs that were further

than 50 words apart). Since all these items are strong associates, one might expect that the word association ranking should correlate with co-occurrence frequency for both HAL's neighbors and non-neighbors (recall that these two groups of words collectively show a +.25 correlation between ranking and co-occurrence). The results were striking. The correlation using the close neighbors is +.48; the correlation for the non-neighbors is +.05. These results suggests that the popular view that association is reflected by word co-occurrence seems to be true *only* for those items that are similar in the first place.

Word association does not seem to be best represented by any simple notion of temporal contiguity (local co-occurrence). From the perspective of the HAL model, word meaning is best characterized by a concatenation of these local co-occurrences, i.e., global co-occurrence -- the range of co-occurrences (or the word's history of co-occurrence) found in the word vector. A simple co-occurrence is probably a better indicator of an episodic relationship, but a poor indicator for more categorical or semantic knowledge. One way to think about global co-occurrence is that it is the contextual history of the word. The weighted co-occurrences are summed indices of the contexts in which a word occurred.

*Lesioning word meaning vectors.* Another way to consider what little effect the local co-occurrence information has on vector similarity is to remove it from the vector and recompute similarity. Consider, for example, the *cat - dog* example. Somewhere in the vector for *cat* there is the vector element that is the weighted local co-occurrence of *cat* when it was preceded by *dog* (matrix row) and the weighted co-occurrence of *cat* when followed by *dog* (matrix column). For any word pair one could remove the vector elements that correspond to the local co-occurrences for those two words. We did this for the prime - target pairs for the stimuli that were used in the semantic priming studies described above (e.g., Lund, et al. 1995, 1996; originally from Chiarello, et al. 1990). There were several items that were not in the HAL lexicon, but this left 286 related prime - target pairs. The procedure resulted in two sets of vectors for these related pairs: an original set with all vector elements and another set in which the elements corresponding to the words themselves had been removed. This lesioning of the vector elements that corresponds to the words themselves removes the effect of their local co-occurrence. The correlation was then computed for the prime - target distances for these two sets of items. There was virtually no impact of the removal of these vector elements (the correlation was 0.99964). This may not seem as

counter-intuitive when one considers that removing the local co-occurrence amounts to the removal of only 1/70,000 of the word's vector elements. What is important is the overall pattern of vector similarity (global co-occurrence), particularly for the rows and columns for which the variance is largest (thus indicating greater contextual exposure).

### Rethinking Syntactic Constraints

That a common word representation can carry information that is both semantic and grammatical raises questions about the potential interaction of these kinds of information and subsequent sentence-level comprehension. Burgess and Lund (1997a) addressed this issue by using the semantic constraint offered by a simple noun phrase on the syntactic processing of reduced-relative sentences. English is a language with a SVO bias (Bever, 1970) where the sentential agent is typically in the subject position (e.g., 1a). Sentence (1a) follows this construction and is simple past tense. Sentence (1b) has the same initial three words, *The man paid*, which might lead the parser to construct a past-tense construction. However, when the preposition, *by*, is encountered, it becomes clear to the comprehension system that the sentence structure is past participle. These reduced-relative past-participle constructions are usually difficult to understand. When the semantics of the initial noun phrase constrain the interpretation, such that the initial noun is not a plausible agent for the verb (as in 1c), reading difficulty can be reduced.

- (1a.) The man paid the parents.
- (1b.) The man paid by the parents was unreasonable.
- (1c.) The ransom paid by the parents was unreasonable.

Although it makes intuitive sense that semantic plausibility facilitates interpretation, an important question in psycholinguistics has been the speed at which this can take place and the implication of processing on architectural modularity. A variety of investigators have shown that semantic plausibility plays an immediate role in the interpretation of these constructions so that syntactic reinterpretation is not necessarily required (Burgess & Hollbach, 1988; Burgess & Lund, 1994; Burgess, Tanenhaus, & Hoffman, 1994; MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Carlson, 1989; Trueswell, Tanenhaus, & Kello, 1993; Trueswell, Tanenhaus, & Garnsey, 1994; see MacDonald, et al. 1994, for a review). Other investigators, however, have found that this type of semantic constraint does not immediately affect this

sentential interpretation and that the reader will always initially misinterpret a construction like (1c) even with the strong semantic constraint (Frazier, 1978; Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983).

Several studies directly compared stimulus sets used by various investigators. Some sets produced results that reflected this initial semantic effect, and other sets did not (Burgess & Lund, 1994; Burgess, et al. 1994; Taraban & McClelland, 1988). These studies have found that the strength of the semantic constraint differed in important ways between some of these experiments and that this difference predicts whether or not the reading difficulty is eliminated.

Burgess and Lund (1997a) pursued this issue of the strength of semantic constraint on syntactic processing by evaluating how well distance in a high-dimensional context space model (HAL) would correspond to the constraint offered by a sentence's initial noun and the past-participle verb (e.g., *man-paid* vs *ransom-paid*). They theorized that the context distance between noun-verb pairs would be inversely correlated with reading ease. Burgess and Lund used context distances for stimuli from three different studies which all used these reduced relative past-participle sentence constructions to simulate the results from these three experiments. One study did not find an effect of this noun context on the reading time in the disambiguating region. The other two studies, which they simulated, did find this context effect, suggesting a more constraining relationship between the biasing noun and the verb. Burgess and Lund's results showed that HAL's context distances were shorter for the stimuli used in the two studies that did find a context effect than for the study that did not find the context effect. Thus, it appears that HAL's representations can be sensitive to this interaction of semantic and grammatical information and that context distance provides a measure of the memory processing that must accompany sentence comprehension. This is because HAL's similarity measure is essentially a measure of contextuality, a notion upon which we will expand later. These results suggest that a high-dimensional memory model such as HAL can encode information that can be relevant beyond just the word level. Based on these kinds of results, we certainly cannot make any general claims about modeling syntax with high-dimensional meaning spaces. At the same time, however, it does seem clear that the distance metric corresponds to constraints between different grammatical classes of words that have specific contextual relationships in sentences. Furthermore, Elman (1990) has shown that sentential meaning can be tracked in an attractor

network (a 70 dimensional space). His results demonstrated that a network can learn grammatical facts about complex sentences (relative clauses; long-distance dependencies). The relationship between what these high-dimensional spaces can represent and their correspondence to higher-level syntactic forms remains an exciting and controversial domain.

### **Rethinking Representational Modularity**

Whether or not the syntactic processor can utilize contextual information to guide its parsing decision has been a controversial issue; the question itself presupposes a parsing mechanism. Recent theories of parsing have been driven by lexical/semantic models of word recognition. The notion of a two-stage parser, where a syntactic structure is built without initial recourse to the available semantics, continues to be a dominant theory in psycholinguistics (Clifton & Ferreira, 1989; Frazier & Clifton, 1996). More recent models of syntactic processing have relied increasingly on the richness of the lexical/semantic system to provide the various semantic, thematic, and local co-occurrence information required to correctly assign meaning to word order (Burgess & Lund, 1994; MacDonald, et al. 1994; Tanenhaus & Carlson, 1989). Basic constraint satisfaction models are free to utilize a broad range of information and further acknowledge that these different sources of information vary in their relative contribution to the sentence comprehension process. The evidence that supports a constraint-satisfaction approach calls into question any strict notion of modularity of processing. Recent results suggest that the language processor is not modular, and that whether or not modular performance is observed is a function of a variety of constraints that may or may not be available.

A parallel issue exists with respect to modularity of representations. Most theories of language comprehension assume that different forms of representations (e.g., syntactic, grammatical, lexical, and semantic) are linguistically distinct, regardless of their position on processing modularity (Burgess, 1998; Burgess & Hollbach, 1988; Burgess & Lund, 1994; Frazier, 1978; Frazier & Fodor, 1978; MacDonald, et al. 1994; Tanenhaus & Carlson, 1989). Connectionist word recognition models have tended to blur this distinction by consolidating the learning from different representational sources into a single layer of hidden units (Elman, 1990; Seidenberg and McClelland, 1989). HAL's vector acquisition process simply accumulates a word's representation from the word's surrounding context. Each vector element for a particular word corresponds to a symbol

(usually another word) in the input stream that was part of the contextual history for that particular word. The word's representation, then, corresponds to the complete contextual learning history that is a function of the word's context, the frequency of co-occurring symbols, and the relative weight in the moving window. Our previous work with semantic priming, word association norms (Lund, et al. 1995, 1996), and other grammatical effects (Burgess & Lund, 1997a; Burgess, Livesay, & Lund, 1998; also see Finch & Chater, 1992) suggest that HAL's representations carry a broad range of information that account for a variety of cognitive phenomena. This generality of HAL's representations suggests that it is possible to encode many "types" of semantic, grammatical, and possibly syntactic information into a single representation and that all this information is contextually driven.

With the increased reliance on contextual factors and their influence in syntactic processing, the need for a representational theory is vital. We propose that the vector representations that are acquired by the HAL model can provide at least a partial resource. These vector representations are a product of considerable language experience (~300 million words of text in these simulations) that reflect the use of words in a highly diverse set of conversational contexts. The model does not presuppose any primitive or defining semantic features, and does not require an experimenter to commit to a particular type or set of features. Rather, the model uses as "features" (i.e., the vector elements) the other words (and symbols) that are used in language. That is, a word is defined by its use in a wide range of contexts.

### Rethinking Similarity

The notion that word meaning and similarity are somehow constrained by the contexts in which they are found is uncontroversial. Many possible relationships between context and word meanings was delineated by Miller and Charles (1991). Their strong contextual hypothesis that "two words are semantically similar to the extent that their contextual representations are similar" (p. 8) seems quite superficially consistent with much of what we have been presenting, and, in many ways, it is. However, Miller and Charles rely heavily on a (commonly held) assumption that we think becomes problematic for a general model of meaning acquisition. It is important for Miller and Charles that similarity is closely attached to grammatical substitutability. Much of HAL's generalizability would be quite limited if the acquisition process somehow hinged on grammatical substitutability.

The context in which a word appears in HAL is the 10 word window that records weighted co-occurrences prior to and after the word in question. However, this local co-occurrence is abstracted immediately into the more global representation. The result is that a word's meaning ultimately has little to do with the words that occur in close temporal proximity to it.

The role of context is transparent in the HAL model. Word meanings arise as a function of the contexts in which the words appear. For example, *cat* and *dog* are similar because they occur in similar sentential contexts. They are not similar because they frequently co-occur (locally). This is a departure from traditional views on similarity which focuses on item similarity. The vector lesioning experiment produced an important insight by simply removing the vector elements that correspond to the locally co-occurring words in a pair of vectors and recomputing their distance in the hyperspace. This manipulation made virtually no difference. Another example that illustrates the lack of effect of local co-occurrence is the relationship between *road* and *street* (see Figure 1). These two words are almost synonymous, however they seldomly locally co-occur. These two words do, however, occur in the same contexts. This lack of an effect from local co-occurrence is also found with Landauer and Dumais' (1997) high-dimensional memory model and would appear to be a general feature of this class of model.

As a result of the role of contextual similarity, words may possess elements of items similarity, but it is due to the role of the context. An advantage to this notion of contextual similarity (rather than the traditional item similarity) is that words that are related in more complex, thematic, ways will have meaningful distance relations. For example, *cop* and *arrested* are not traditionally "similar" items. However, they are contextually similar, and as a result, the distance between such items reflects the relationship between the agent and action aspects of the lexical entities (see Burgess & Lund, 1997a). This greatly expands the potential role of similarity in memory and language models that incorporate meaning vectors such as these.

### A Comparison of Dynamic Learning Models

Although it is argued that HAL is a dynamic concept acquisition model (the matrix representing a momentary slice in time), the prototypical "dynamic learning model" is probably the more established connectionist model. In this section, we will compare the output of the global co-occurrence learning algorithm with a SRN (simple recurrent network) when both are given the same input corpus. The motivation for this comparison is that both claim to be models that learn from context. HAL uses a weighted 10-word moving window to capture the context that surrounds a word. The example SRN used for this comparison is that of Elman (1990) in which the context for a target word in a sentence is the recurrent layer that provides an additional set of inputs from the previous word to the hidden units encoding the current word whose representation is being learned. HAL and this SRN also have in common that the words are represented in a distributed fashion and in a high-dimensional meaning space. The meaning space in HAL is the 140,000 elements defined by the input symbols that are weighted by the global co-occurrence procedure. The meaning space in Elman's SRN is a function of the hidden unit activations.

Elman (1990) used a SRN which was trained to predict upcoming words in a corpus. When the network was trained, hidden unit activation values for each input word are used as word representations. The corpus he used was one constructed from a small grammar (16 sentence frames) and lexicon (29 words); the grammar was used to construct a set of two and three word sentences resulting in a corpus of ~29,000 words. The corpus is simply a sequence of words without sentence boundary markers or punctuation. This corpus was fed into a neural network consisting of input, hidden, and output layers plus a fourth context layer which echoed the hidden layer (see Figure 3). The network was trained to predict the next word, given the current word and whatever historical information was contained in the context layer. At the end of training, the hidden layer activation values for each word were taken as word representations.

Our approach to replicating this used the global co-occurrence learning algorithm in the HAL model. A co-occurrence matrix was constructed for the Elman (1990) corpus using a window size of one word. As the context represented in Elman's neural network consisted of only prior items, word vectors were extracted from the co-occurrence matrix using only matrix rows (representing prior co-occurrence), yielding twenty-nine vectors of twenty-nine elements

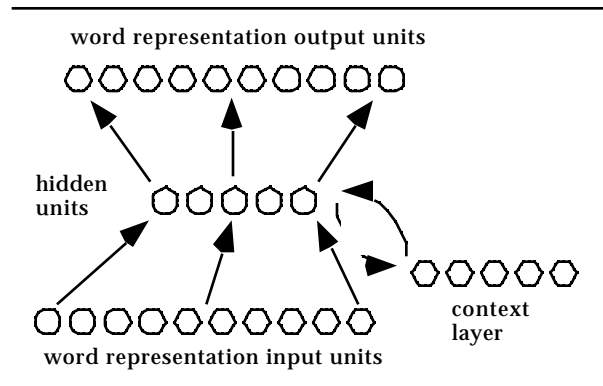


Figure 3. Elman's (1990) simple recurrent neural network architecture.

each. These vectors were normalized to constant length in order to account for varying word frequency in the corpus.

A gray-scaled representation of the co-occurrence matrix for the 29 lexical items is shown in Figure 4. In this figure, darker cells represent larger co-occurrence values, with rows storing information on preceding co-occurrence and columns following co-occurrence. For example, the matrix shows that the word *glass<sub>row</sub>* was often preceded by the words *smash* and *break*; *eat<sub>column</sub>* was often followed by all of the animates except for *lion*, *dragon*, and *monster* (who were presumably the agents involved). A casual examination of this matrix suggests that semantic information has been captured, as words with similar meanings can be seen to have similar vectors. To more closely examine the structure of these vectors,

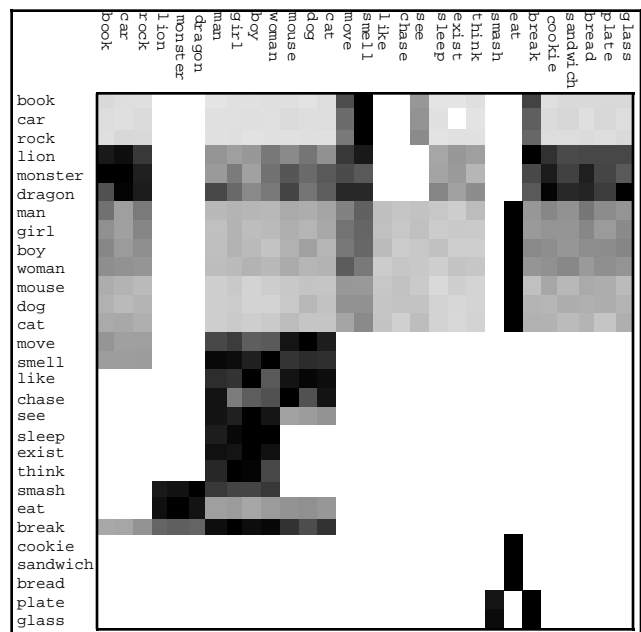


Figure 4. Gray-scaled representation of the global co-occurrence matrix for the 29 lexical items used in Elman (1990).

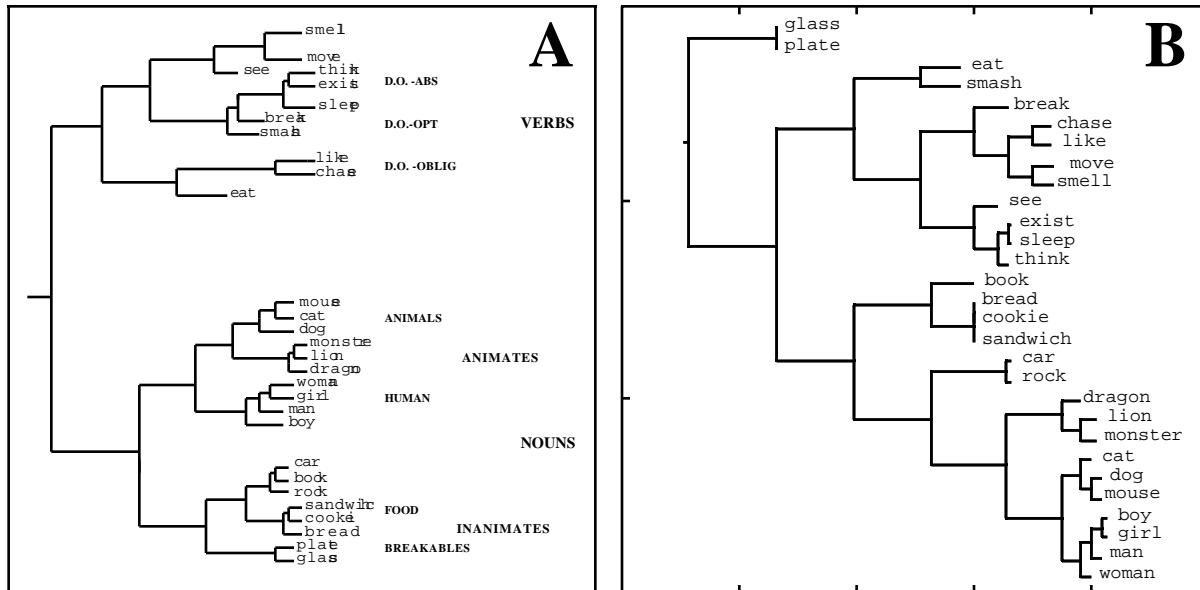


Figure 5. Hierarchical cluster diagrams of (A) Elman's (1990) results with hidden unit activation vectors from a simple recurrent neural network, and, (B) results using the global co-occurrence vectors from HAL model trained on Elman's corpus.

we constructed a hierarchical clustering of HAL's vectors, shown in Figure 5b, alongside the clustering obtained by Elman (1990) in Figure 5a. HAL performed reasonably in separating animate objects (with subdivisions for people, dangerous animals, and safe animals), edible objects, verbs, and fragile objects. These categorizations are similar to those found by Elman. Although the clustering produced by HAL is not as clean as that of Elman's SRN, it should be noted again that the matrix was formed with a very conservative one-word window, while the test sentences were often three words long.

Similar results were produced by two approaches to the generation of semantic structure. Why should such apparently dissimilar approaches yield the same results? The answer is that both techniques capitalize on the similarity of context between semantically and/or grammatically similar words in order to construct representations of their meanings. Virtually the only thing that the two approaches have in common, in fact, is that they both have context information available to them. That they both find the same basic underlying structure within the vocabulary argues strongly that context is a valid and fundamental carrier of information pertaining to word meaning. The SRN appears to be a little more sensitive to grammatical nuances. It also produces more compact representations, as the vectors are shorter than the vocabulary size (one element per hidden unit). However, it has a drawback in that it does not scale well to real-world vocabularies. If tens of thousands of words are to be tracked (not just 29),

not only would the network be enormous, but training it would be difficult and time consuming (if not just impossible) due to the sparseness of the representations to be learned. It is important to know that global co-occurrence models yield virtually the same result as the SRN. The equivalence of these two approaches should facilitate the understanding of the general role of context as well as beginning to develop hybrid models.

### The Symbol Grounding Problem

Glenberg (1997) raises two issues that he claims are serious problems for most memory models. First is the symbol-grounding problem. The representations in a memory model do not have any extension to the real world. That is, lexical items can not be understood with respect to just other lexical items. There also has to be a grounding of the representation of the lexical item to its physical reality in the environment (cf., Cummins, 1996). A model that represents a concept by a vector of arbitrary binary features or by some set of intuitively reasonable, but contrived, set of semantic features does not have a clear mapping onto the environment that it supposes to represent. HAL takes a very different approach to this problem. In HAL, each vector element is a coordinate in high-dimensional space for a word. What is important to realize about each vector element is that the element *is* a direct extension to the learning environment. A word's vector element represents the weighted (by frequency) value of the relationship between the part of the environment

represented by that element and the word's meaning. The word's meaning is comprised of the complete vector. Symbol grounding is typically not considered a problem for abstract concepts. Abstract representations, if memory models have them, have no grounding in the environment. Again, though, HAL is different in this regard. An advantage to the representational methodology used in HAL is that abstract representations are encoded in the same way as more concrete words. The language environment, i.e., the incoming symbol stream, that HAL uses as input is special in this way. That is, abstract concepts are, in a sense, grounded. The second problem faced by models that develop "meaningless" internal representations is that the variety of input that a human can experience does not get encoded, and, therefore, the memory representation is inevitably impoverished. With the current implementation of HAL, this is certainly a limitation. The learning experience is limited to a corpus of text. This raises an important but currently unanswerable question. Are the limitations in HAL's representations due to the impoverished input or will higher-level symbolic representations be required to flesh out a complete memory system as argued by Markman and Dietrich (in press)? We do think a HAL-like model that was sensitive to the same co-occurrences in the natural environment as a human-language learner (i.e., a model that is completely symbol grounded, using more than just the language stream) would be able to capitalize on this additional information and construct more meaningful representations. Any answer to these questions would be premature and speculative. That said, however, these are important issues for a general model, and we will present what we think are intriguing (although speculative) arguments that high-dimensional memory models can capture some aspects of schemata and decision making.

### Higher-level Cognition

We have previously argued that HAL's word vectors generated by the global co-occurrence learning mechanism are best regarded as encoding the information that will model the initial bottom-up activation of meaning in memory. Semantic and grammatical structure emerge from what we refer to as global co-occurrence which is the (weighted) concatenation of thousands of simple, local co-occurrences or associations. However, others have maintained that statistical associations are unlikely to produce sophisticated knowledge structures because they do not encode the richness of the organisms interaction with the environment (Glenberg, 1997; Lakoff, 1991; Perfetti, 1998). Lakoff argues that

schemata are a major organizing feature of the cognitive system and that the origin of primary schemata involve the embodiment of basic sensory-motor experience. Glenberg takes a similar stance concluding that complex problem solving is beyond the scope of simple associationist models. Although simple association can be part of some similarity judgments (Bassok & Medin, 1997), Gentner and Markman (1997) maintain that higher-level structure is typically involved in making similarity judgments. It is easy to imagine that a model such as HAL would be less than adequate for representing higher-level cognition. Markman and Dietrich (in press) suggest that the adequacy of a cognitive model will require multiple grainsizes. Symbolic representations may not represent the fine grain necessary for context sensitivity. Conversely, distributed representations are limited in how they can manage contextual invariance. In this section, we address how high-dimensional memory models may offer a plausible representational account of schematic representations and some forms of decision making.

### Rethinking Schemata

A schema is typically considered a symbolically structured knowledge representation that characterizes general knowledge about a situation (Schank & Abelson, 1977). Schemata can be instantiated in distributed representations as well (Rumelhart, Smolensky, McClelland, & Hinton, 1987). Rumelhart, et al. modeled the notion of "rooms" by having a set of microfeatures that corresponded to aspects of various rooms (e.g., television, oven, dresser, etc). Each of these microfeatures can fill a slot in a schema. The primary difference between a symbolic account and a distributed account is that in the distributed account the schema is not a structured representation -- a distributed schema is a function of connection strengths.

In HAL, the notion of a schema best corresponds to the context neighborhood. A word in HAL's lexicon can be isolated in the high-dimensional space. Surrounding this word will be other words that vary in distance from it. Neighbors are words that are close. Table 3 shows the context neighborhoods for three words (*beatles*, *frightened*, and *prison*). An MDS solution can demonstrate that different sets of words can be plausibly categorized. It remains unclear exactly what the space in an MDS figure represents. The context neighborhood provides more of an insight into the nature of the meaningful information in the hyperspace. A schema is more specifically structured than a context neighborhood. What they both have in common is that components of a

schema or context neighbors of a word provide a set of constraints for retrieval. The context neighborhoods are sufficiently salient as to allow humans to generate the word from which the neighbors were generated or a word closely related to it (Burgess, et al. 1998). The neighborhoods provide a connotative definition or schema of sorts; not the denotative definition one would find in a dictionary.

One criticism of spatial models such as HAL is that the words in the meaning space have sense, but no reference (Glenberg, 1997). This is generally true; many models have features that are provided by intuition, or are hand coded, or derived from word norms. As a result, there is no actual correspondence between real input in a learning environment and the ultimate representations. There are several models, including HAL, that differ in this regard (also see Landauer & Dumais, 1997, LSA model; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990, and Elman's, 1990, connectionist approach to word meaning). In other words, they are symbol grounded with respect to the environment that serves as input (a stream of language in these cases). Edelman (1995) has taken a similar approach to constructing representations of the visual environment.

Another criticism of high-dimensional space models is that they do not adequately distinguish between words that are synonyms and words that are antonyms (Markman and Dietrich, in press). This can be illustrated by the neighbors of *good* and *bad*. *Bad*'s closest neighbor is *good*. Such examples also highlight the difference between item similarity and context similarity and is usually seen with adjectives. *Good* and *bad* occur in similar contexts (good and bad are in the eye of the beholder) and will tend to be close in meaning space. Spatial models will tend to have this problem. Although this is a limitation, it may not be as problematic as suggested by Markman and Dietrich. *Good*'s immediate neighbors contain more items related to its core meaning (*nice, great, wonderful, better*) than items related to *bad*. Likewise, *bad*'s neighbors share its meaning (*hard, dumb, stupid, cheap, horrible*) more so than *good*'s meaning. Despite this limitation, we would argue that the neighborhoods offer sufficient constraint to characterize meaning.

### Problem Solving

Problem solving and decision making are complex cognitive events, both representationally and from the view of processing. It would be premature indeed to suggest that high-dimensional memory models can purport to model the range of representations that must provide the scaffolding for complex problem

Table 3. Nearest neighbors for *Beatles*, *frightened*, and *prison*.

Beatles	Frightened	Prison
original	scared	custody
band	upset	silence
song	shy	camp
movie	embarrassed	court
album	anxious	jail
songs	worried	public

solving. High-dimensional memory models may, however, be useful in modeling aspects of problem solving that hinge on similarity. For example, Tversky and Kahneman's (1974) approach to decision making about uncertain events relies on representativeness and availability. In HAL, representativeness might be captured by context similarity. Likewise, a frequency metric is likely to predict availability. Tversky's (1977) feature contrast model has been used to model many kinds of similarity judgments. Prior to Tversky, similarity relations were, for the most part, considered to be symmetric. Tversky has shown that it more likely that asymmetry is the rule. An example from Tversky illustrates this: North Korea is judged to be more similar to China than China is to North Korea. Featural asymmetry is now acknowledged to be an important component of many models of similarity (Gentner & Markman, 1994; Medin, Goldstone, & Gentner, 1993; Nosofsky, 1991) and models of metaphor (Glucksberg & Keysar, 1990).

The metric from the HAL model that is typically used is the distance metric. For example, *tiger* and *leopard* are 401 RCUs apart in the high-dimensional space. This is useful information; we know that *tiger* and *leopard* are more contextually similar than *tiger* and *bunny* or *eagle*. However, context distance is symmetrical and this would seem to be an important limitation of HAL. Others have noted that the areas around items in a high-dimensional space can vary in density (Krumhansl, 1978; Nosofsky, 1991). HAL is no different. Tversky (1977) pointed out how *tiger* is a more probable response to *leopard* in a word association task than *leopard* is to *tiger*. Although *tiger* and *leopard* are 401 units apart; their context neighborhoods differ in the items they contain and in their density. In HAL's high-dimensional space, *tiger* is the 4th neighbor to *leopard*, whereas *leopard* is the 1335th neighbor to *tiger* -- an asymmetry in the direction one would find with word norms (see Figure

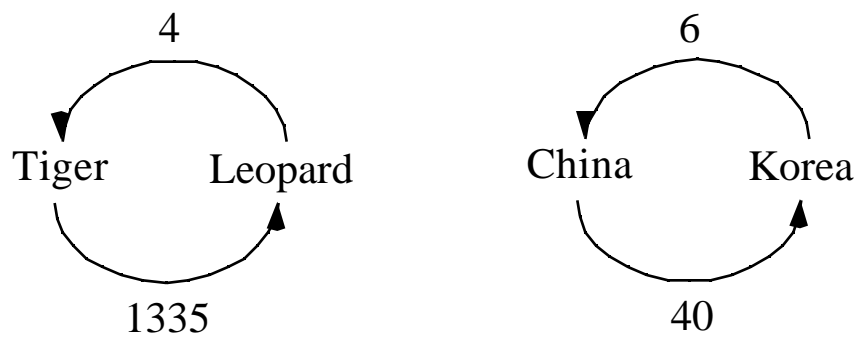


Figure 6. Diagram illustrating the asymmetry in the number of context neighbors separating two word pairs.

6). The Korea - China example from Tversky shows a similar asymmetry in the number of intervening neighbors (see Figure 6). China is Korea's 6th neighbor in HALs hyperspace; Korea is China's 40th neighbor. Density in the HAL model can also be an important metric in predicting semantic effects. Buchanan, Burgess, and Lund (1996) found that context density was a better predictor of semantic paralexias with brain-damaged patients than was either context distance or word association norm rankings.

The characteristics of context neighborhoods (density and neighbor asymmetries) would seem to be important factors in modeling the representations important to the problem solving process. This is not to say that distance is not. The ability to use similarity information in sorting is an important ability. Tversky and Gati (1978) had subjects select one country from a set of three that was most similar to a comparison target. They found that frequently the choice hinged on the similarity of the other two possible choices. Before simulating the similarity component among country choices and the implication of similarity in the sorting task, it was important to show that HAL's vector representations could reflect the semantic characteristics of geographic locations. To do this, names of cities, states, and countries were submitted to a MDS procedure (see Figure 7). The figure reflects how the vectors could be used to categorize locations. It is important to note that English speaking countries seem separated in this space from Asian countries since our analysis of the Tversky and Gati sorting experiment requires that the vector representations for countries reflect semantic distinctions of countries.<sup>3</sup>

In this demonstration, we attempted to show that low-level contextual information can support high-level decision making. In the Tversky and Gati (1978) experiment subjects were asked to match a country to the most similar of three other countries. One of the three countries varied, with the assumption that changing the third choice country could affect which of the other two countries would be chosen as the closest match to the comparison target. Indeed, the manipulation of the third choice country tended to cause a reversal in which of the other two choices were chosen as the closest match to the comparison target.

Following Tversky and Gati (1978), the assumption was made that subjects were not actually finding the closest match to the target country (if they had been, there would have been no reversal), but that instead they were finding the pair among the choices which were most similar, and then assigning the remaining item as the closest match to the comparison target. This is a rather simplistic model, disregarding the target item, but in theory it can account for the choice reversal found by Tversky and Gati.

To evaluate this theory context distances were computed for each triplet of choices (Tversky & Gati, 1978, Table 4.4). Note that no distances were computed that related to the target item. The two countries which had the smallest context distance between them were considered to form their own match, with the third country then being considered to be matched to the target. For an example of this procedure, see Figure 8. Here it was predicted that, in

<sup>3</sup> Proper name semantics have a tradition of being notorious difficult to model (see Burgess & Conley, in press-a,b). The simulation of the Tversky and Gati

(1978) experiment with HAL vector representations is notable in that it represents another successful application of the model to the implementation of proper name semantics.

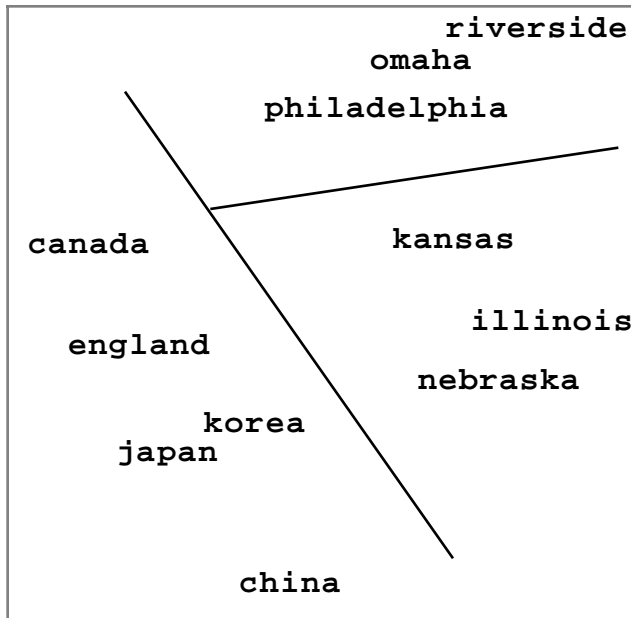


Figure 7. Two-dimensional multidimensional scaling solutions for countries, cities, and states.

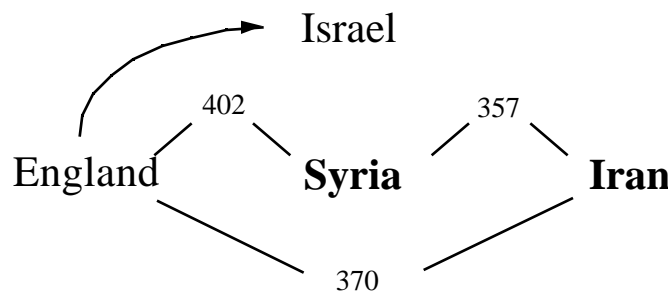
set 1, Israel would be matched with England because Syria and Iran will tend to form their own grouping. When Syria is replaced by France, in set 2, the prediction is that France will now tend to pair up with England, leaving Iran as the best match to Israel. In fact, this reversal was found by Tversky and Gati 61.2% of the time.

This result is well replicated using HAL

distances. In set 1, Syria and Iran do indeed have the smallest inter-item distance (357 RCUs), leading to England being paired with Israel. And in set 2, England and France are very similar in HAL's hyperspace (distance of 230 RCUs), which brings about the same result as found in humans: Israel is now paired with Iran. Across all of Tversky's stimuli, the analysis using context distances from the HAL model led to the expected country being picked 62% of the time, a very similar result to the human data.

The contrast model of Tversky and Gati's (1978) computes similarity between items as a combination of their common and distinctive features. With humans the sorting task requires that attention be directed to common features of the choices with the result that these features become more salient. A different choice option redirects attention to other common features resulting in the different pairing. HAL is a representational model, and, as implemented, does not have a mechanism that corresponds to attention. Consequently, these results suggest that the contextual information available in HAL's vector representations is sufficient for this type of decision making. It should be emphasized that we are not claiming that HAL is a decision-making model. Rather, we feel that a contextual model of meaning can provide sufficiently rich information about concepts such that this information can be useful in higher-level decision making.

Set 1



Set 2

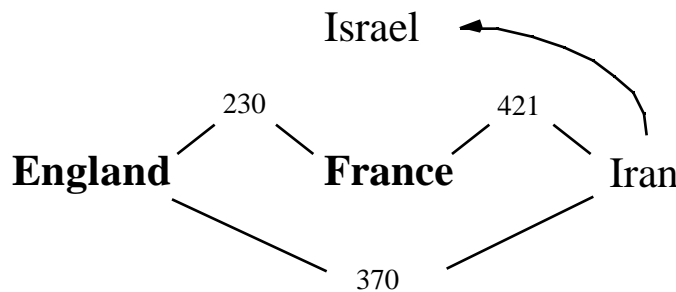


Figure 8. An example of two sets of countries from Tversky and Gati (1978). Context distances (in RCUs) are indicated for all possible pairs of choice words. Word pairs with the shortest distance are in bold.

## What are HAL's Vector Representations?

"Form and function are one."

Frank Lloyd Wright

The inner workings of many models are rather opaque. Shepard (1988) characterized this criticism well with connectionist models: "... even if a connectionist system manifests intelligent behavior, it provides no understanding of the mind because its workings remain as inscrutable as those of the mind itself" (p. 52). It is difficult at times to understand the precise representational nature of hidden units or the psychological reality of "cleanup" nodes. Conversely, the HAL model is quite transparent, and it is also quite simple. One goal of the HAL project has been to "do much, with little," to the extent possible. For some, these features are problematic because they can not possibly capture the "... range of human abilities that center on the representation of non-co-occurring units, especially in language" (p. 12) (Perfetti, 1998). Addressing the question of what HAL's vector representations represent involves a number of subtle descriptive and theoretical issues.

*What are HAL's vector representations:*

*A descriptive answer.*

The meaning vector is a concatenation of local co-occurrences in the 10-word window. The first time two words co-occur, there is an episodic trace. An example in Table 1 would be a co-occurrence value of 5 for *raced* as preceded by *horse*. This 5 represents a strong episodic relationship in the window between *horse* and *raced*; it is a strong relationship since the words occurred adjacently. As soon as *raced* co-occurs with some other word this cell in the matrix starts to lose its episodic nature. As experience accrues, the vector elements acquire the contextual history of the words they correspond to. The more a word experiences other words, the richer its context vector. Although a complete vector has the 70,000 row elements and the 70,000 column elements, approximately 100 - 200 of the most variant vector elements provide the bulk of the meaning information. The vector (whether full or a smaller set of features) is referred to as a global co-occurrence vector; local co-occurrence is simply a co-occurrence of one item with another.

*What are HAL's vector representations:*

*A theoretical answer.*

There are probably three ways in which one can coherently consider what these vector representations are. The representations are of words, thus, in a sense

they are symbolic. The vectors also have some of the characteristics of distributed representations as well. Finally, one could consider the vector representations simply a documentation of the learning history of a word in many contexts and not worry about mediationism and representation. Each of these possibilities will be briefly addressed to answer the question "what are HAL's vector representations?"

*Vectors as symbolic representations.* Each element in a vector representation provides one coordinate in the high-dimensional space. The vector provides a set of coordinates or constraints that converge on a symbol (usually a word). In the hyperspace, one can retrieve a word's neighbors which are usually other words. In addition, each vector element directly corresponds to a symbol in the input stream. Of course, this is a result of using text as input. There is evidence, however, that suggests that a process similar to global co-occurrence can deal with the speech segmentation problem at the phonetic level (Cairns, Shillcock, Chater, Levy, 1997). One can imagine a global co-occurrence system that operates at two cascaded levels in which the speech segmentation processor could present the meaning processor with its output.

*Vectors as distributed representations.* The meaning of a word is a function of a pattern of values (of the different vector elements). A word is a point in the hyperspace, but this point is the convergence of thousands of flexible constraints. The memory matrix is a slice in time of the history of the system as it encounters language experience. HAL's vectors have several important characteristics of distributed representations. They degrade gracefully. This is clear in the extreme when one considers that several hundred of the 140,000 vector elements will suffice for most purposes. Another characteristic of distributed representations is that they are comprised of subconceptual elements. A typical connectionist example is how *dog* and *cat* might have subconceptual features such as <has-legs> and <does-run> (from Hinton & Shallice, 1991). Presumably these features are some of the perceptual components from which the concept will develop. HAL's representations are acquired from the language environment and the representations take on a more abstract form than a set of concrete objects. The "subconceptual" features in HAL are not just other symbols, but the weighted co-occurrence value with the other symbol. In a sense, calling word features subconceptual is misleading -- it depends on how and where in the nervous system the perceptual apparatus parses the input. What is probably more important at this point in the development of the theory of

meaning is the notion that concepts are comprised of a large set of co-occurrence elements. These co-occurrence values form the contextual history of a word. Thus, *dog* and *cat* are similar in HAL because they occur in similar contexts, not because they both are furry, small, have four legs, and are pets (although these features constrain their appearance in particular contexts). As a result, the degree to which items locally co-occur is of very little relevance in the development of the distributed meaning vector. Recall the experiment in which the locally co-occurring vector elements were lesioned. When similarity was recomputed, the effect was negligible. This runs counter to many uses of co-occurrence in memory models (Perfetti, 1998). HAL's distributed vector representations are representations of contextual meaning (very much like LSA, Landauer & Dumais, 1997).

The representation of meaning in a high-dimensional space means that there are parallels to earlier high-dimensional models of similarity (Osgood, et al. 1957; Shepard, 1988; Smith, et al. 1974; Tversky, 1977). An important difference is that HAL is also an acquisition model that relies on context, not human similarity judgments or normative data, for its derivation of meaning.

*Vectors as representations of learning history.*

Associationist theory holds at its core the notion of temporal contiguity. In the HAL model, temporal contiguity is closely related to local co-occurrence. At the risk of being redundant, contextual similarity is a function of global co-occurrence, not local co-occurrence. Each vector element is one of many measures of a word's experience in the context of another word. As Deese (1965) pointed out over 30 years ago, the basic principles of association are best viewed in the context of distributions of associations to particular events or stimuli. It is these higher-order associations (or global co-occurrence) that reveal the structure in memory and language. Simply incorporating first-order association (temporal contiguity) into a memory model is an invitation to either underestimate the effect of association or to set up a strawperson model. Classical and sometimes instrumental learning principles have found a home in connectionist models and are certain to in high-dimensional memory models that essentially instantiate very high-order association to build semantic and grammatical structure. Viewing meaning vectors as a learning history would seem to obviate the need for representations, per se. As such, a disadvantage to this view is the discomfort it will generate in legions of cognitive scientists. Giving up mediationism forces a theorist into the realm of

functionalism. This approach focuses on the relationship of the learning environment and the contextual history of the learner. The failure of associationist models to have a more influential role in the last 30 years may hinge on their reliance on word association methodologies, moreso than theoretical limitations. Current high-dimensional memory models can simulate the acquisition process using substantial amounts of experience in order to model the psychological plausibility of a range of cognitive phenomena. The closer relationship between the actual learning environment, context, and vector behavior may reduce the need for a reliance on a host of memory metaphors currently employed in cognitive science. Such a view of HAL's representations are likely to be viewed as radical (Markman & Dietrich, in press) or terribly misguided (Glenberg, 1997). However, Watkins (1990) has argued that it is a mistake to try and justify complex models because one is trying to model complex phenomena. Regardless, global co-occurrence offers a very principled approach to developing structured representations from real environment.

It may be premature to decide upon the ultimate veracity of these views of HAL's representations. These three views all have some relevance to high-dimensional memory models -- at a minimum we hope that this state will facilitate further discussion about the nature of high-dimensional representations.

## Conclusions

The notion of similarity can be found in many psychological models of memory and language. In high-dimensional memory models such as HAL (Burgess, 1998; Burgess & Lund, 1997a; b; Lund & Burgess, 1996), LSA (Foltz, 1996; Landauer & Dumais, 1997), or other similar approaches (Bullinaria & Huckle, 1996; Finch & Chater, 1992; Schutze, 1992), the conceptual representations are a product of the contexts in which words are found. The HAL model is distinguished by a number of very simple assumptions about how concepts are acquired. Despite these limitations (or perhaps because of them), the range of cognitive phenomena that the model has been applied spans from basic word recognition and meaning retrieval (Lund, et al. 1995, 1996), semantic dyslexia (Buchanan, et al. 1996), grammatical effects (Burgess & Lund, 1997a), abstract meaning and emotional connotation (Burgess & Lund, 1997b) and sentence and discourse comprehension (Burgess, et al. 1998; also see Foltz, 1996, & Landauer & Dumais, 1997). Most of the work with the HAL model has focused on the nature of representations rather than on processing issues.

The memory matrix is a slice in time of the concept acquisition process. An advantage to this is that representational issues can be explored independently of processing constraints. The drawback, of course, is that one can not evaluate the interaction of the two. There are currently two exceptions to this in the research with the HAL model. The process of acquisition as presented in this chapter affords a look at a number of important issues such as the role of associations in the learning process and how categorical knowledge is formed from these simpler constructs (Burgess, et al. 1997). HAL's representations have been incorporated in a mathematical memory processing model of hemispheric asymmetries (Burgess & Lund, 1998). Furthermore, the results from the global co-occurrence mechanism compare favorably with neural net implementations as presented earlier.

The ability to separate in a computational model the representational and the processing components and to provide a set of real-valued meaning vectors to the process provides the initiative to begin rethinking a host of important issues such as the nature of similarity, representational modularity, and how a

computational model can have its representations grounded in its environment. The HAL model is proposed as a model of the initial bottom-up component of meaning activation. Higher-level meaning and problem solving may not be beyond the scope of the model as previously thought. Despite the range of problems that the HAL model has been applied to, there are many unanswered and exciting questions. One of the most important is the extent to which global co-occurrence and distributed representations can account for higher-level cognition as the model is expanded to encounter more of a plausible environment beyond just language input. It does, however, seem very clear that HAL's focus on context has been very beneficial and is likely to continue to provide insights into the contextually dynamic form of mental representations and their role in cognitive processing.

#### References

- Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory & Language*, 36, 311-336.
- Berwick, R. C. (1989). Learning word meanings from examples. In D. L. Waltz (Ed.), *Semantic structures: Advances in natural language processing* (pp. 89-124). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. ). New York: John Wiley and Sons.
- Buchanan, L., Burgess, C., & Lund, K. (1996). Overcrowding in semantic neighborhoods: Modeling deep dyslexia. *Brain and Cognition*, 32, 111-114.
- Bullinaria, J. A., & Huckle, C. C. (1996). *Modelling lexical decision using corpus derived semantic representations in a connectionist network*. Unpublished manuscript.
- Burgess, C. (1998). From Simple Associations to the Building Blocks of Language: Modeling Meaning in Memory with the HAL Model. *Behavior Research Methods, Instruments, and Computers*, 30, 1-11.
- Burgess, C., & Conley, P. (in press-a). Developing a semantics of proper names. *Proceedings of the Cognitive Science Society* (pp. xx - xx). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Burgess, C., & Conley, P. (in press-b). Representing Proper Names and Objects in a Common Semantic Space: A Computational Model. *Brain & Cognition*.
- Burgess, C., & Hollbach, S.C. (1988). A computational model of syntactic ambiguity as a lexical process. In *Proceedings of the Tenth Annual Cognitive Science Society Meeting* (pp. 263-269). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in Context Space: Words, Sentences, Discourse. *Discourse Processes*, 25, 211-257.
- Burgess, C., & Lund, K. (1994). Multiple constraints in syntactic ambiguity resolution: A connectionist account of psycholinguistic data. *Proceedings of the Cognitive Science Society* (pp. 90-95). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Burgess, C., & Lund, K. (1998). Modeling cerebral asymmetries of semantic memory using high-dimensional semantic space. In Beeman, M., & Chiarello, C. (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (p. 215 - 244). Hillsdale, N.J.:

- Lawrence Erlbaum Associates, Inc.
- Burgess, C., & Lund, K. (1997a). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210.
- Burgess, C., & Lund, K. (1997b). Representing abstract words and emotional connotation in high-dimensional memory space. *Proceedings of the Cognitive Science Society* (pp. 61-66). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Burgess, C., Lund, K., & Kromsky, A. (1997, November). *Examining issues in developmental psycholinguistics with a high-dimensional memory model*. Paper presented at the Psychonomics Society Meeting, Philadelphia, PA.
- Burgess, C., Tanenhaus, M. K., & Hoffman, M. (1994). Parafoveal and semantic effects on syntactic ambiguity resolution. *Proceedings of the Cognitive Science Society* (pp. 96-99). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33, 111-153.
- 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 and Language*, 38, 75-104.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The M.I.T. Press: Cambridge, MA.
- Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language & Cognitive Processes*, 4, 77-103.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York, N.Y.: Academic Press.
- Cottrell, G. W. (1988). A model of lexical access of ambiguous words. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution in the comprehension of human language* (pp. 179-194). Los Altos, CA: Morgan Kaufmann Publishers.
- Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA, MIT Press.
- Cushman, L., Burgess, C., & Maxfield, L. (1993, February). *Semantic priming effects in patients with left neglect*. Paper presented at the International Neuropsychological Society, Galveston, TX.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, 391-407.
- Deese, J. (1965). *The structure of associations in language and thought* (pp. 97-119). Baltimore: The Johns Hopkins Press.
- Dyer, M. G. (1990). Distributed symbol formation and processing in connectionist networks. *Journal of Experimental and Theoretical Artificial Intelligence*, 2, 215-239.
- Edelman, S. (1995). Representation of similarity in 3D object discrimination. *Neural Computation*, 7, 407-422.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Finch, S., & Chater, N. (1992). Bootstrapping syntactic categories by unsupervised learning. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (820-825). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, 5, 335-339.
- Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments & Computers*.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. Thesis, University of Connecticut. Indiana University Linguistics Club.
- Frazier & Clifton, (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291-325.
- Gallant, S. I. (1991). A practical approach for representing context and for performing word sense disambiguation using neural networks. *Neural Computation*, 3, 293-309.
- Gentner, D., & Markman, A. B. (1997). The effects of alignability on memory. *Psychological Science*, 8, 363-367.
- Gernsbacher, M. A. (1990). *Language comprehension*

- as structure building. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Glenberg, A. M. (1997). What memory is for. *Behavioral & Brain Sciences*, 20, 1-55
- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3-18
- Hinton, G. E., McClelland, J. L. & Rumelhart, D. E., (1986). Distributed representations. In Rumelhart, McClelland, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 1: Foundations* (pp. 77-109). Cambridge: MIT Press.
- Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74-95.
- Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, 112, 500-526.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.
- Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1, 97-120.
- Lakoff, G. (1991). Cognitive semantics.
- Landauer, T. K., & Dumais, S. (1994, November). *Memory model reads encyclopedia, passes vocabulary test*. Paper presented at the Psychonomics Society.
- 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 Bulletin*, 104, 211-240.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, 203-208.
- Lund, K., Burgess, C., & Atchley, R.A. (1995). Semantic and associative priming in high-dimensional semantic space. *Proceedings of the Cognitive Science Society* (pp. 660-665). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Lund, K., Burgess, C., & Audet, C.. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. *Proceedings of the Cognitive Science Society* (pp. 603-608). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning & Verbal Behavior*, 23, 709-733.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- Markman, A., & Dietrich, E. (in press). In defense of representation.
- Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 3-23.
- Maxwell, A. (1976). *The Singer enigma*. New York, NY.: Popular Library.
- McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In Rumelhart, McClelland, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2: Psychological and biological models* (pp. 272-325). Cambridge: MIT Press.
- McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 558-572.
- McRae, K., de Sa, V., & Seidenberg, M. S. (1996). The role of correlated properties in computing lexical concepts. *Journal of Experimental Psychology: General*, 126, 99-130.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- Miller, G. (1969). The organization of lexical memory: Are word associations sufficient? In G. A. Talland & N. C. Waugh (Eds.), *The pathology of memory* (pp. 223-237). New York: Academic Press.
- Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6, 1-28.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140.
- Osgood, C. E. (1941). Ease of individual judgment-processes in relation to polarization of attitudes in

- the culture. *The Journal of Social Psychology*, 14, 403-418.
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, 49, 197-237.
- Osgood, C. E. (1971). Exploration in semantic space: A personal diary. *Journal of Social Issues*, 27, 5-64.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, University of Illinois Press.
- Palermo, D. S., & Jenkins, J. J. (1964). *Word association norms grade school through college*. Minneapolis, MN: University of Minnesota Press.
- Perfetti, C. A. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, 25, x-y.
- Plaut, D. C., & Shallice, T. (1994). *Connectionist modelling in cognitive neuropsychology: A case study*. Hove, England: Lawrence Erlbaum Associates, Inc.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, 12, 1-20.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1987). Schemata and sequential thought processes in PDP models. In Rumelhart, McClelland, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2: Psychological and biological models* (pp. 7-57). Cambridge: MIT Press.
- Schutze, H. (1992). Dimensions of meaning. In *Proceedings of Supercomputing '92* (pp. 787-796). New York: Association for Computing Machinery.
- Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organizations*. Norwood, N.J: Ablex Pub. Corp.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1191-1210.
- Shepard, R. N. (1988). How fully should connectionism be activated? Two sources of excitation and one of inhibition. *Behavioral and Brain Sciences*, 11, 52.
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts, Inc.: New York.
- 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.
- Tanenhaus, M. K., & Carlson, G. N. (1989). Lexical structure and language comprehension. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 529-561). Cambridge, MA: MIT Press.
- Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role expectations. *Journal of Memory and Language*, 27, 597-632.
- Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318.
- Trueswell, J.C., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528-553.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Tversky, A., & Gati, (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 79-98). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Watkins, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist*, 45, 328-3.