INVITED ADDRESS

From simple associations to the building blocks of language:
Modeling meaning in memory with the HAL model

CURT BURGESS
University of California, Riverside, California

This paper presents a theoretical approach of how simple, episodic associations are transduced into semantic and grammatical categorical knowledge. The approach is implemented in the hyperospace analogue to language (HAL) model of memory, which uses a simple global co-occurrence learning algorithm to encode the context in which words occur. This encoding is the basis for the formation of meaning representations in a high-dimensional context space. Results are presented, and the argument is made that this simple process can ultimately provide the language-comprehension system with semantic and grammatical information required in the comprehension process.

"...meaning" seems to constitute, for most psychologists at least, something inherently nonmaterial, more akin to "idea" and "soul" than to observable stimulus and response, and therefore to be treated like the other "ghosts" that J. B. Watson dispelled from psychology. (Osgood, Suci, & Tannenbaum, 1957, p. 1)

There are many ways to characterize meaning (Ogden & Richards, 1923). Lexical semantics usually refers to the meanings of individual words and distinctions constrained by their morphology. Structural semantics involves the integration of these word meanings into a more complex meaning guided by the syntax of the sentence. Of course, we do not tend to remember sentences; we abstract this level of meaning to what can be referred to as a mental model of a situation in which this linguistic information—lexical-semantic, structural, and/or our experience of the environment—combine to form a complex mental representation that is not limited to the linguistic form (see Garrett, 1986; Johnson-Laird, 1983). The purpose of this paper is to describe a new class of memory models that are referred to as high-dimensional memory models. These models have as their goal extracting and representing meaning from a stream of language. The nature of the meaning that can be extracted with these models is at the level of the word and to some extent, relevant to larger units of meaning. Although computational high-dimensional memory models can be considered "new," the fact is that they have many connections to past and present modeling efforts. (Chase & Engle, 1973; Osgood et al., 1957) should be considered the father of high-dimensional memory models. These models also have at their heart a clear associationist bent (Deese, 1965). Another more recent functional precursor to high-dimensional memory models is Elman's (1990) work with recurrent neural networks that learn generalized representations of their experience with input from the environment. Elman's work has operationally defined the surface context is important.

At the same time, however, there are limitations to these historical precursors that set the stage for the advantages of high-dimensional memory models. Osgood et al.'s (1957) semantic differential approach required vast numbers of human judgments about the nature of words in order to derive a set of semantic dimensions. In addition, the semantic differential technique (although very useful) requires one to commit to a set of semantic features upon which all words will literally be judged. This is a very tricky enterprise, which makes very strong presumptions about how the language user interacts with the environment and encodes experience. Deese's (1965) work hinges on the use of word association norms that correspond, in some sense, to meaning to provide a set of interrelations among words. Both word-association norms and the semantic differential technique require considerable human overhead to gather word-meaning information for even a small set of items. It is also the case that neither of these approaches offers any model of how information gets organized in the first place. In contrast, connectionist learning models have a completely different type of limitation. It would simply take an unreasonable amount of computer-processing time to develop a model of memory from the combination of a very large corpus and a set of lexical items used to support human language. The time results from the complexity of the learning algorithm and the number of passes through a corpus that a connectionist model requires to complete its learning. Fortunately, at a functional level, we have demonstrated that recurrent neural networks and our global co-occurrence learning procedures are both learning models that utilize a word's context to induce word meaning and, given the same input, will produce very similar results (Lund & Burgess, 1997).

This paper describes a particular high-dimensional memory model, the hyperospace analogue to language (HAL) model. The model makes completely explicit the nature of associations and the relationship of associations to categorical knowledge. As a result, this model can address a variety of difficult and problematic theoretical issues in learning and memory. This will occur in the context of describing how global lexical co-occurrence (the learning procedure that HAL uses) can provide some of the basic building blocks of a language-comprehension system.

THE HAL MEMORY MODEL

Having a plausible methodology for representing word meaning is an important component to a model of memory, particularly when the goal is to extend the model to sublexical processing. HAL uses a large corpus of text, and the basic methodology involves tracking lexical co-occurrences within a 10-word moving window that "slides" along the text. Using these co-occurrences, a high-dimensional meaning space of 140,000 lexical dimensions can be developed. Since each vector element represents a symbol (usually a word) in the text input, we refer to this high-dimensional meaning space as a context space.

This high-dimensional context (or meaning) space is the memory matrix that can be used to simulate experiments or to further analyze word meanings.

Constructing the Memory Matrix

As mentioned before, the basic methodology of the model involves developing a matrix of word co-occurrence values for a set of words by moving a 10-word window along the corpus of text. Within the moving window, word-by-word co-occurrence counts are tabulated and are inversely proportional to the number of words separating a pair of words. Words that are closer together in the moving window get a larger weight. Table 1 shows an example of the syntax matrix for selecting a five-word window using the example sentence "the horse raced past the barn fell." The matrix is actually two triangular matrices folded into one table—that is, rows encode co-occurrence values that occur before a word, and columns encode co-occurrence information occurring after the word. Consider, for example, the word barn in this sentence. Co-occurrence values of words occurring prior to barn are in the barn row. The word "past" is separated by one word from barn and thus gets a 4. If past had occurred adjacent to barn, it would have received a 5. There are two occurrences of the word the in the example sentence, and there is a 6 recorded for barn the word. One occurrence of the occurs just prior to barn and gets a 5. The other the in this sentence is five words away and gets a 1. The 6 in this cell represents the addition of the five and one co-occurrence values. Columns work the same way (to encode the subsequent co-occurrences), except that the point of reference is from the column word. This example was made using a five-word window; however, a 10-word window was used in the actual model, and HAL's matrix is 70,000 square items.

The corpus that served as input to the HAL model is approximately 300 million words of English text gathered from Usenet newsgroups that contained English text. Properties of Usenet text that were appealing were both its conversational nature and its diverse nature, making it closer in form to everyday speech. An important goal in this project is to develop a model that does minimal preprocessing of the input. It is to HAL's credit that it works with this noisy, conversational input, thus managing some of the same problems that the human-language comprehender encounters.

Matrix Properties

HAL's vocabulary consists of the 70,000 most frequently used words in the corpus. About half of these symbols had entries in the standard Unix dictionary. The remaining items were nonword symbols, misspellings, proper names, and slang. These atypical items were kept in the lexicon since large representations proclivity. Each row in the matrix represents the degree to which each word in the vocabulary precedes the word corresponding to the row. Similarly, each column represents the co-occurrence values for words following the word corresponding to the column.

| Table 1 Sample Global Co-Occurrence Matrix for the Sentence "the horse raced past the barn fell"
<table>
<thead>
<tr>
<th>barn</th>
<th>horse</th>
<th>past</th>
<th>raced</th>
<th>the</th>
</tr>
</thead>
<tbody>
<tr>
<td>barn</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>horse</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>past</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>raced</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>the</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Note — The values in the matrix rows represent co-occurrence values for words that preceded the word (row label). The values in the columns represent co-occurrence values for words following the word (column label). Cells containing zeros were left empty in this table. This example uses a five-word co-occurrence window.
Vector Properties

A full co-occurrence vector for a word consists of a concatenation of the row and the column for that word. Thus, the vector for *barn* (from Table 1) would be 0, 2, 1, \(3, 0, 6, 0, 0, 0, 0, 0\). The vector for *barn* in the full matrix would be 140,000 elements long rather than 10 elements long as in this example.

We view these vectors as coordinates of points in a high-dimensional space, with each word occupying one point. Each vector corresponds to a lexical symbol in the input stream. These symbols are not always words (they may be numbers, emotions, etc.), although it is convenient to refer to them as word vectors. In most simulations, vectors of 140,000 elements in length are used; however, we should not conclude that we think it best to characterize human memory as represented by 140,000 dimensions. We suspect that much less dimensionality is required. For our work, only 100–200 vector elements seem necessary. Some effects can be easily carried by as few as 10 vector elements, which is useful in connectionist models where the training time is often non-linearly related to vector size. What is important in any attempt to ‘horton’ a vector is to keep the vector elements that are most informative.’ To determine this, the column and row variance is calculated, and the elements with the smallest variance are not used. Variance as an information measure is useful in a model like HAL since it is a direct reflection of the diversity of the contexts in which words are found. An element with only 100–200 vector elements contain most of the variance in these word vectors. For example, a 200-element word vector would not use the lowest variance 139,800 vector elements. In most of our simulations, however, we use the full vector simply because it is computationally more straightforward than computing variance across such a large matrix of numbers.

Using this type of vector representation, differences between semantic and phonological vectors (e.g., *france* and *panama*) can be measured as the distance between high-dimensional points defined by their vectors. Distance between two words can be computed using a similarity metric. We typically use a Minkowski metric (see Eq. (1)) and usually use a Euclidean distance metric (\(r = 2\)).

\[
\text{distance} = \sqrt{\sum (x_i - y_i)^2}, \quad (1)
\]

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 (not just item) similarity. The vectors can also be viewed graphically as can be seen in Figure 1. Sample words (e.g., *france, puppy*) are shown with their accompanying 25-element vectors (the 25 most variant elements). Each element has a continuous numeric value (the frequency normalized value from its matrix cell). A gray scale is used to represent the normalized value, with white 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 (e.g., 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 (Figure 1). See Lund and Burgess (1996) for a full description of the HAL methodology.

The 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 their ‘features’ other words (symbols in the input stream) that contribute to the ability to link the representations for abstract concepts (e.g., *justice, reality*) as easily as one can have representations for more basic concepts (e.g., *dog, book*) (Burgess & Lund, 1997b). This is important if not absolutely crucial, when developing a memory model that purports to be general in nature and that provides the essential bottom-up input to a language-comprehension system.

**CONSTRUCTING THE BUILDING BLOCKS OF LANGUAGE**

The goal of language understanding is to derive meaning from sounds. To build a language comprehension model, one has to describe how word-level categorical information is formed. The HAL model encodes categorical information, both semantic and grammatical. These two sources of information provide the front end to the ability of combining words into larger meaningful units.

**Semantic Knowledge**

There are numerous models of semantic memory, and, as with other aspects of cognition, considerable disagreement over details. However, there are several commonly held assumptions. Semantic models tend to focus on words and their relationship to other words, sense rather than reference. These relationships also tend to be summed across experience, rather than representing specific episodes. At the same time, these semantic memories are constructed as a function of our episodic experience with the world, and they allow for the categorization of information.

Similarly, there are many ways in which cognitive psychologists can assess the structure and processing components of semantic memory. There are three ways in which we have investigated semantic structure with the HAL model, often in conjunction with experiments with human subjects. First, HALS word vectors seem to possess sufficient information from the global co-occurrence procedure to support basic semantic categorization. Second, we have compared distances in the high-dimensional memory space with the priming data that can be obtained by using the semantic priming methodology. This work has implications for how we think about word associations. Lastly, the area in the high-dimensional space surrounding a word contains other words that constitute a context (or a semantic) neighborhood for that word. We will review evidence that suggests that HALS context neighborhoods possess certain denotive characteristics of word meaning.

Semantic categorization. Using human judgments about concepts to determine categorical structure has a long history in cognitive psychology. In their well-known study, Rips, Shoben, and Smith (1973) used typological ratings of different kinds of birds to generate a set of prototypes for the semantic verification model. They then used a multidimensional scaling (MDS) procedure to transform these ratings into a two-dimensional pictorial representation of these different types of birds, which demonstrated that the typicality ratings provided important information for categorization purposes. With HAL, rather than human ratings, the word vectors can be used to categorize concepts. The vectors represent the contextual history of a word, and we have seen that the distance between these vectors is meaningful from the standpoint of word neighborhoods and semantic priming. In Figure 1, we see that categorical structure emerges when we view the MDS solution for body parts, animals, and geographic locations. These three groups of words segregate nicely. It is important to note that all the MDS presentations presented in this paper are a reduction of 140,000 dimensions to two dimensions, which results in a loss of resolution in the categorical integrity that can be conveyed. Still, it is clear that the word vectors allow for categorization much like human ratings of similarity. The objects in this MDS are all concrete nouns. Other categorization work has shown that HALS representations extend to the categorization of abstract nouns and emotional words (Burgess & Lund, 1997b).

**Semantic neighborhoods**

The meaning space in HAL is a high-dimensional space in which each word is represented as a point in space along a slice of the vector coordinates. The distances that these words are from one another are results of the contexts in which they systematically occurred. Similar words occur in similar contexts. As a result, one should be able to see semantic and contextual relationships among the words that are close to another word in the hyperspace. We have found that these neighborhoods are rich in meaningfully related concepts. Table 2 shows the six closest neighbors for two words: *beetles* and *frightened*. Inspection suggests that the notion of a musical group is inherent in the neighborhood for *beetles*. Other neighbors of *beetles* further constrain the concept (best, british, greatest). The nature of these neighborhoods is more connotative than denotive. Some of *beetles* neighbors are concepts that are definitional (i.e., *beetle* is a kind, is likely considered by many as original, and has produced many songs and albums (and movies)). The semantic neighborhood is more of a definition by implication. It certainly is not a typical denotive definition such as "English quartet of composers and musicians; members are .." (Morris, 1971). The neighbors represent a set of constraints on meaning. Some neighbors may be almost synonymous; other neighbors characterize various aspects of meaning. Meaningful neighborhoods are not limited to nouns. Table 2 also contains the neighborhood for *frightened* and is similarly informative providing synonyms and words related to other emotional aspects of *frightened*. These neighbors are the closest set of words to fright- ened. The nature of the meaning acquisition process sug-
Table 2

Example Neighborhoods for 
beauties and frightened

don't

Word
beauties
original
long
music
almond
songs
scared
unlighten
shy
embarrassed
anxious
worried

Table 3

Example Prime-Target Word Pairs From the Semantic, 
Associated, and Semantic + Associated Relatedness Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Word Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic</td>
<td>table-bed, music-art</td>
</tr>
<tr>
<td>Associated</td>
<td>cradle-baby, mag-beer, mud-tired</td>
</tr>
<tr>
<td>Semantic + Associated</td>
<td>ale-beer, uncle-unt, half-be</td>
</tr>
</tbody>
</table>

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

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. One set of stimuli that we have evaluated in detail using relatedness judgments is the study by Shenton 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 prime-target pairings were strongly related by using a categorically (e.g., road-street, girl-boy) (see Shenton et al., 1995). Using HAL, we argued that the semantic-only condition did not produce priming simply because the prime-target pairings 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 stimuli 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 (in preparation) shows that our HAL simulation that Shenton and Martin's (1992) failure to find priming was due to insufficient relatedness in their semantic-only condition. Recall that they used an unprimed control list (Richards et al. 1997) in the one of the earliest findings showing that the strength of association did not correlate with priming. Similarly, Chiarello, Burgess, Richards, and Pollock (1990) found semantic-only priming using a low pair of related trials and a naming task. However, Lupker (1984) did not find priming for semantically related word pairs that were not also associated related. A similar set of results is found in Shelton and Martin (1992). They used a single presentation of 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 relation. It is not clear what kind of semantic priming under these conditions. A comparison of experiments such as these usually entails a comparison of the methodologies. Experiments that do not observe semantic-only priming typically avoid the lexical decision task, unless it is part of the individual presentation procedure (i.e., Shelton & Martin, 1992). The naming task is thought to be less sensitive to strategic effects (although this may also limit its sensitivity to semantic relations). Clearly, experimental procedures and task differences play a part in these results. Focusing on task differences, however, may divert attention from important representational properties of the stimuli to be just as important. In developing representational theory, it is important not to make representational conclusions based solely on procedural issues.
relationships that is produced when a person free associates. Yet this is an unsatisfying definition at a theoretical level. It also confounds many closely associated pairs that can be found using a word-association procedure. One intuitive conceit of word association is that it is related to the degree to which words tend to co-occur in language (Miller, 1965). Spence and Owens (1980) confirmed this long-held belief empirically. To see whether this relationship between word association ranking and lexical co-occurrence held for the language corpus that we use for HAL, we compiled 380 highly associated pairs from the Palermo and Jenkins (1964) norms as the basis for the experiment (Lund et al., 1996). We replicated Spence and Owens’s 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 nonneighbors (pairs that were further than 50 words apart). Since all strong items are strong associations, one might expect that the word-association ranking should correlate with co-occurrence frequency for both HAL’s neighbors and HAL’s nonneighbors (recall that these two groups were, words collectively, positively correlated between ranking and co-occurrence). The results were striking. The correlation using the close neighbors is +.48; the correlation for the nonneighbors is +.05. These results suggest that the popular view that association is reflected by word co-occurrence seems to be true only for items that are similar in the first place. Word association does not seem to be best represented by any simple notion of the 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. Next, we will review evidence that the contextual nature of the origin of meaning in HAL’s representations provide more than semantic knowledge.

Grammatical Knowledge

In the previous section, the high-dimensional meaning spaces were described as both semantic and contextual. The bulk of the early work with the HAL model investigated issues that were typically semantic in nature: various types of semantic and associative priming, use of neighborhoods as word definitions, semantic paralexias, and accounting for results of word production norms. The simple idea behind the HAL model (and certainly not unique to our work) that context determines word meaning has been presented in another linguistic domain. Ervin (1963; Ervin-Tripp, 1970) and many others (see Nelson, 1977) have shown that a child’s experience with context is implicated in the acquisition of both semantic and grammatical knowledge. There is compelling evidence that the word vectors encoded in HAL’s global co-occurrence procedure carry information about grammatical class. Perhaps it is not surprising that the contextual nature of the representations could encode grammatical information, given the importance of context in a developing child’s grammatical competence. At the same time, however, a singular acquisition mechanism that can learn semantic as well as grammatical-class information violates the traditionally held notion of representational modularity (see Burgess & Lund, 1997a). We have explored these grammatical and syntactic representational issues in detail elsewhere (Burgess et al., 1995; Burgess & Lund, 1997a; see Finch & Chater, 1992, for a similar approach).

In this section, two analyses of grammatical class will be presented. One shows that word vector representations encode grammatical-class information as well as semantic information. Vectors of words of different grammatical classes were extracted from the model and analyzed using an MDS procedure, just as with the semantic categorization presented above. In Figure 3, it can be seen that the nouns, verbs, determiners, and prepositions cluster into their own spaces. Words that are noun–verb ambiguous (sketch, building) straddle the high-dimensional fence between these two grammatical categories. The same pattern of grammatical-class discrimination was seen in Burgess and Lund (1997a), in which a more systematic approach was taken. Nouns and verbs that had been part of the stimuli from an online parsing experiment were used, along with a larger set of determiners and prepositions (Simulation 2). The results presented here replicate the basic grammatical categorization effect found in the Burgess and Lund (1997a) paper.

Nouns, verbs, prepositions, and determiners occur in very different parts of a sentence due to their differing roles in the syntactic structure of the sentence. From the perspective of contextual constraints and how substitutability will play a role in these contextual constraints, it follows that these coarsely coded grammatical categories would be part of what is encoded in a global co-occurrence procedure such as HAL’s. A much more challenging problem would be the categorization of subclasses of classes all belonging to the same grammatical category. Figure 4 illustrates this with different types of determiners. Articles (a, an, the), demonstratives (this, these, those), and generics (her, his, our, their) all separate into their own portion of the context space. This result goes beyond the earlier result that demonstrated the separation of the distinct grammatical classes. In this case, these determiners occupy the same location in the surface structure of a sentence: a dog, this dog, her dog. However, the higher order, distributional information associated with these different classes of determiners has to be implicit in the word vectors in order to obtain a result such as that shown in Figure 4. A similar pattern can be seen when we look at the high-dimensional spaces occupied by simple past participles and past-tense verbs, as seen in Figures 5A and 5B, respectively (modeled after

Burgess & Lund, 1997a, Simulation 3). These verbs are unambiguous and clearly distinguished by their vector representations. The exception to this is the past participle verb seen. Although seen is a past participle, it is completely contextually uninformative, almost grammatical; it is essentially used as a past tense (as in “I seen it”), which is reflected in the input corpus and, ultimately, in the vector’s word representation and its migration to the past-tense space. In Burgess et al. (1998), an analogous pattern of results was seen with verbs of implicit causality, such as flatter and praise. Verbs of implicit causality are interpersonal verbs that have a semantic bias that specifies whether the agent or the patient is most likely to carry out the action of the verb. This becomes clear in Sentences 1a and 1b.

1a. The sportscaster praised the Huskers.
1b. The sportscaster flattered the Wolverines.

In Sentence 1a, the sportscaster is praising a team because of their impressive victory in the Orange Bowl. In Sentence 1b, the verb flattered suggests an ulterior motive on the part of the sportscaster—probably a disingenuous attempt to interview players after a narrow victory in the Rose Bowl. Au (1996) has subjects sort verbs of implicit causality into “meaningful” groups. The MDS of the human categorization reflecting Au’s results and the MDS of these verbs in Burgess et al. (1998) are very similar. The overall grammatical distinction is made in both results, and internal semantic relationships are also present.

Grammatical information seems to be carried in HAL’s word vectors. This can be seen for grammatical categories that vary greatly in their syntactic role (nouns, verbs, prepositions, and determiners). It can also be seen for more subtle distinctions within grammatical classes (determin-
rs, different types of verbs). The analyses presented here (and elsewhere; see Burgess & Lund, 1997a) suggest that it is useful to distinguish between declarative (class- and semantic information) and propositional (class- and semantic information) in a sense, the world's learning history. This historical history (referred to as global co-occurrence) was shown to be informative in modeling a variety of lexical and semantic effects. Two domains—semantic and grammatical categorization—were considered. We saw that HAL can be used to replicate the semantic priming effect by use of the distance metric. One of the first HAL results used the semantic distance metric to analyze two different sets of prime-target pairs from two different experiments. The analysis of Shelton and Martin's (1992) stimuli indicated that there was an asymmetry in the semantic relatedness that might have resulted in an overestimation of the effect. More generally, it was found that the correlation between word-association ranking and (local) word co-occurrence was a product of the contextual similarity of the words. Words that are not similar in a local context are not likely to be associated in a global sense.

Rethinking Similarity

Most notions of similarity hinge on the idea that concepts share some set of features (Kornatka, 1992). A car and a truck may seem similar because they both have tires, can be steered, transport people and things from place to place, and so on. These shared features provide the mechanism for similarity judgments, word recognition, or use by higher level comprehension systems. The notion of similarity in the HAL model is more complex than in previous models such as the semantic similarity space. Although there is evidence that suggests that semantic similarity does not produce automatic bottom-up priming (Lupker, 1984; Shelton & Martin, 1992), recent results with visual-semantic patients (Cutshall et al., 1993) and with the same procedures used by Shelton and Martin (i.e., McRae & Boisvert, in press) converge on the conclusion that, with sufficiently similar word pairs, retrieval is rapid and automatic.

The priming results suggest that there is a meaningful organization of the spaces around a word. Neighborhoods of words close to a target word in the hypherspace mimic what could be viewed as the semantic organization that our brain uses to represent words. The close neighbors to beatles reveal that the word has relationships with original and band, and to the concepts song and album. More impressively, human subjects are able to use the words in the neighborhood to converge on the word that produced the neighborhood in the first place. An important limitation of these neighbor-


NOTES

1. I do not want to leave the impression that there is just one high-dimensional computational memory model. Kevin Land and I started working on the HAL (Hyperspace Analogue to Language) model in 1992. The first conference presentation discussing HAL was in 1994, and the first publication was in 1995. The work of Landauer and Dumais (1997) and earlier work (also see Foltz, 1996) is very closely related to our model. In fact, they are very close cousins—at a general level, the basic approach is the same. They differ in a number of implementational aspects, some of which may suggest important distinctions. Also, Nick Chater's (Finch & Chater, 1992) work takes a similar approach, although Chater is more reluctant to see these models as psychological models.

2. The process of frequency normalization and computing the distance metric are intertwined. The distance metric in HAL is referred to as a Euclidean distance metric. This is an arbitrary but frequency-normalized Euclidean distance metric. The first step in normalizing a vector is to compute its magnitude (sum the squares of the elements, take the square root of that, and divide by the number of elements). The magnitude is then divided by 666.0, and each vector element is divided by the resulting number.

3. Visual inspection of the MDS presentations in this paper appears to show a robust separation of the various word groups. However, it is important to determine whether these groupings are clearly distinguished in the high-dimensional space. Our approach then is to use an analysis of variance that compares the individual distances within the groupings. This approach is accomplished by calculating all combinations of item-pair distances within a group and comparing them with 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.

(Manuscript received January 16, 1998; revision accepted for publication March 5, 1998.)