

High-dimensional semantic space accounts of priming [☆]

Michael N. Jones ^{a,*}, Walter Kintsch ^a, Douglas J.K. Mewhort ^b

^a *Institute of Cognitive Science, University of Colorado, 344 UCB, Boulder, CO 80309-0344, USA*

^b *Department of Psychology, Queen's University, Kingston, Ont., Canada K7L 3N6*

Received 2 March 2006; revision received 26 June 2006

Available online 18 September 2006

Abstract

A broad range of priming data has been used to explore the structure of semantic memory and to test between models of word representation. In this paper, we examine the computational mechanisms required to learn distributed semantic representations for words directly from unsupervised experience with language. To best account for the variety of priming data, we introduce a holographic model of the lexicon that learns word meaning and order information from experience with a large text corpus. Both context and order information are learned into the same composite representation by simple summation and convolution mechanisms (cf. Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626). We compare the similarity structure of representations learned by the holographic model, Latent Semantic Analysis (LSA; Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–240), and the Hyperspace Analogue to Language (HAL; Lund, K., & Burgess, C., (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, 203–208) at predicting human data in a variety of semantic, associated, and mediated priming experiments. We found that both word context and word order information are necessary to account for trends in the human data. The representations learned from the holographic system incorporate both types of structure, and are shown to account for priming phenomena across several tasks.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Semantic space models; Models of word meaning; Semantic priming; Mediated priming

Introduction

A common finding in the psycholinguistic literature is that a word is processed more efficiently when it is preceded by processing of a related word. The common assumption is that the first word (the *prime*) facilitates processing of the second word (the *target*) because it contains within it some of the mental code required for the second response (Rosch, 1975). In semantic priming, the magnitude of facilitation depends on the semantic similarity between the prime and target. For example, *nurse* is processed more efficiently when pre-

[☆] This research was supported by grants from NSERC and Sun Microsystems to DM, and an IERI grant to WK. MJ was supported by a postdoctoral fellowship from NSERC. We would like to thank Mark Steyvers and Jim Neely for comments on an earlier version of this manuscript.

* Corresponding author. Present address: Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405-7007, USA. Fax: +1 812 855 4691.

E-mail address: jonesmn@indiana.edu (M.N. Jones).

ceded by *doctor* than when preceded by *bread* (Meyer & Schvaneveldt, 1971). For this reason, priming has been the predominant task used to study the structure of semantic memory (more specifically, representation of word meaning).

It remains the topic of considerable debate whether semantic priming effects are a result of semantic overlap per se, or are simply due to learned association strength between primes and targets (for reviews, see Huchinson, 2003; Lucas, 2000; McNamara, 2005; Neely, 1991). The debate has important consequences for the opposing localist and distributed approaches to representing word meaning.

Localist models (e.g., semantic networks; Collins & Quillian, 1972) assume that words are represented by nodes of interconnected concepts. Words that are connected to one another by more (or shorter) pathways are more similar in meaning. Localist models account for semantic priming by applying the construct of spreading activation (Collins & Loftus, 1975). When nodes in a network are activated, the activation spreads along the associated pathways to related nodes. The spread of activation makes the connected nodes already partially activated when a related concept is processed. Although spreading activation is an important explanatory concept in semantic networks (Balota & Lorch, 1986), it is important to note that it is a *process* construct that operates on the *structural* representation in a semantic network. In any model, priming requires both an account of the process as well as an account of the structure upon which the process operates.

By contrast, distributed models assume that word meaning is a pattern of elements in an array; the elements may be individually interpretable (e.g., feature lists) or only meaningful as part of an aggregate abstract pattern (e.g., connectionist representations). In a feature list theory (Smith, Shoben, & Rips, 1974), words are represented by lists of binary descriptive features. For example, birds have wings and dogs do not. Semantic priming is accounted for in feature lists simply by overlapping features between the prime and target. Whereas *robin* shares no features with *chair*, it has more shared features with *bat*, and even more with *sparrow*. In a connectionist representation, a word's meaning is distributed over an aggregate pattern of element weights, but none of the elements has interpretable meaning on its own.

A major problem with both feature list and semantic network theories is that the models do not actually learn anything—the semantic representations must be built into the model by the theorist himself. Hand-coded representations rely on intuition of semantic similarity and dimensionality (either by the theorist, or subjective norms, e.g., McRae, de Sa, & Seidenberg, 1997), and may be an inaccurate representation of the information

that is truly salient for semantic representation. Hummel and Holyoak (2003) have noted that hand-coded representations are a serious problem if cognitive modeling is to be a truly scientific enterprise: “All models are sensitive to their representation, so the choice of representation is among the most powerful wildcards at the modeler's disposal” (p. 247).

In addition, hand coding representations artificially hardwires complexity into a model. Assuming that the complexity required for semantic representation is available in the environment, it is more appealing for a model to use simple mechanisms to learn its representations from statistical redundancies in the environment, rather than the theorist building complexity into the model based on intuition. The notion of automatically learning representations from environmental redundancies is the motivation behind recent *co-occurrence models* (e.g., Landauer & Dumais, 1997; Lund & Burgess, 1996).

Co-occurrence models attempt to build semantic representations for words directly from statistical co-occurrences in text. Typically, words are represented in a high-dimensional semantic space (cf. Osgood, 1952, 1971; Salton, 1973; Salton, Wong, & Yang, 1975). For this reason, such models are often referred to as “semantic space” models. Co-occurrence models capitalize on the frequency of words in contexts across a large sample of text. The co-occurrence approach minimizes representation and processing assumptions because much of the model's complexity is learned from the environment—it is not hardwired into the model. For example, to know what *hammer* means, the model will observe all the contexts in which *hammer* is used. One may infer that *hammer* is related to other frequent words in those contexts, such as *nail* and *board*. Further, one may induce that *hammer* is similar to words that appear in similar contexts (i.e., with the same words), such as *mallet* or *screw-driver*. By the same logic, *hammer* is likely less similar to *chromosome* because they tend not to appear in the same or similar contexts.

In Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer & Dumais, 1997), a large-scale text corpus is first transformed into a sparse word-by-document frequency matrix, typically using about 90,000 words and about 40,000 documents. The entries are then converted to log-frequency values, and are divided by the word's entropy, $-\sum p \log p$, over all its documents. Next, the dimensionality of the word-by-document matrix is reduced using singular value decomposition (SVD) so that each word is now represented by a dense vector of approximately 300 dimensions, however, the dimensions have no particular meaning or direct correspondence to the text. SVD has the effect of bringing out latent semantic relationships between words, even if they have never co-occurred in the same document. The basic

premise in LSA is that the aggregate contexts in which a word does and does not appear provides a set of mutual constraints to induce the word's meaning (Landauer, Foltz, & Laham, 1998).

LSA has been successful at simulating a wide range of psychological and psycholinguistic phenomena, from judgments of semantic similarity (Landauer & Dumais, 1997) to word categorization (Laham, 2000), discourse comprehension (Kintsch, 1998), and judgments of essay quality (Landauer, Laham, Rehder, & Schreiner, 1997). LSA has even earned college entrance-level grades on the TOEFL, and has been shown to acquire vocabulary at a rate that is comparable to standard developmental trends (Landauer & Dumais, 1997).

LSA capitalizes on a word's contextual co-occurrence, but not how a word is used in that context. Information about the meaning of *hammer* can be determined by observing the contexts in which it appears. However, the contexts also contain temporal redundancy (grammatical information) about how the word is used relative to other words. Very rarely is a nail ever used to pound a board into a hammer; temporal redundancy reveals information about the word's order relation to other words in the context. Further, this order information reveals that a *hammer* may be more similar to a *mallet* or *hatchet* in how it is used in context than it is to *screw-driver* or *nail*. Even though *screwdriver* and *nail* may have more contextual overlap with *hammer*, they are not used in the same way within those contexts. How a word is used in context can carry as much variance to induce its meaning as what contexts it appears in (and, obviously, these are correlated sources of information).

The Hyperspace Analogue to Language (HAL; Burgess & Lund, 2000; Lund & Burgess, 1996) is related to LSA, but also capitalizes on positional similarities between words across contexts. HAL is trained by moving an n -word window across text and calculating the distance (in word steps) between all words that occur in the window at each point in time. HAL's co-occurrence matrix is a sparse word-by-word ($70,000 \times 70,000$) matrix in which a word's row entry records the frequency, inversely weighted by distance (summed word steps), that the word appeared in the window succeeding every other word possible, and a word's column entry records the frequency (inversely weighted by distance) that the word appeared in the window preceding every other word. After training, the row and column vectors for a word are concatenated to yield the word's representation. Words that have appeared similar distances around the same words can develop similar patterns of elements in their vectors. Thus, both contextual co-occurrence and positional similarity are represented in HAL.

In HAL, words that appear in similar positions around the same words tend to develop the most similar vector representations. Note that two words need not

directly co-occur within the window to develop similar representations. Two words that appeared around the same words will be similar, and this relationship is magnified if they are also found similar distances relative to other words. In HAL, not only do similar nouns (e.g., *cat-dog*) have similar vector representations, but so do other lexical classes, such as determiners, prepositions, and animate and inanimate nouns (Audet & Burgess, 1998; Burgess & Lund, 2000). HAL can be envisioned as a large-scale approximation of the structure that could be learned by a simple-recurrent network (SRN; Elman, 1990, 1991; Servan-Schreiber, Cleeremans, & McClelland, 1991), and has been shown to learn representations that have very similar structure to SRNs when both are trained on small finite-state grammars (Burgess & Lund, 2000). Although HAL does not explicitly encode the order of words, its distance weighting can serve as a proxy for order information (Perfetti, 1998).

LSA and HAL consider subtly different types of information while learning text, and these differences are reflected in the structural representations formed by each model. LSA tends to weight associative relationships more highly than purely semantic relationships. For example, the representation for *car* is much more similar to the representation to *drive* ($\cos = 0.73$) than it is to members of the same semantic category, such as *truck* ($\cos = 0.49$) or *boat* ($\cos = 0.03$). Further, the verb *drive* is more similar to *car* than it is to other action verbs, such as *walk* ($\cos = 0.23$).

By contrast, HAL considers distance between intervening words in the moving window; hence, semantic relationships can become more highly weighted in HAL than associative relationships. In HAL, *car* is more similar to *truck* ($d = 0.90$) and *boat* ($d = 0.95$) than it is to *drive* ($d = 1.12$), and the verb *drive* is more similar to another action verb like *walk* ($d = 1.03$) than it is to *car* ($d = 1.12$).

HAL and LSA focus on different sources of information and, thus, make different predictions about the strength of semantic and associative relationships in memory. The two types of information are correlated, but each model also learns unique variance not considered by the other. The question is whether the unique sources of variance from each type of learning are both needed to account for the structure of semantic memory. Clearly, humans take advantage of both types of information (e.g., the words *hammer* is found in context with and how *hammer* is used relative to those words), and an ideal model of semantic representation would consider both sources of information when learning text.

Attempts to consider both types of information have traditionally used vectors to represent contextual semantics, and rules or productions systems for order information (e.g., Wiemer-Hastings, 2000, 2001). Hence, the two types of information are stored separately and in a different form. Another approach, taken by Dennis

(2005), has used a Bayesian adaptation of string edit theory to represent both syntagmatic and paradigmatic information within a single model. Similarly, Griffiths, Steyvers, Blei, and Tenenbaum (2005) have successfully combined the two sources of information in a generative framework, using a hidden-Markov model to learn sequential dependencies and a probabilistic topic model to learn semantic relationships. Our goal is to apply mechanisms from associative memory theory to learn a single vector representation for a word, containing a mixture of both contextual and word-order information. In doing so, we wish to demonstrate that information about word order is used in representing a word's meaning, and that the simple mechanisms used in other types of associative learning are sufficient to capitalize on this structure without postulating mechanisms for encoding order that are specific to language.

In the domain of associative memory, Murdock (1982, 1992, 1993) has used *convolution* as a mechanism to build associations between pairs of vectors representing words or objects. Murdock represents information about items as random vectors, and information about their associations as convolutions of the item vectors. Both item and associative representations are then summed together and stored in a composite distributed memory representation. The composite representation can be used to determine if an item was learned: A novel item vector will have an expected dot product of zero with the composite representation, and a learned item vector will have a much higher dot product (however, the magnitude depends both on dimensionality and on number of items stored). Further, when a learned item vector is *correlated* with the memory representation (the inverse of convolution), the result is a facsimile of the vector representing the item with which it was associated. If a novel item vector is correlated with the memory representation, the result will not resemble any item known. Murdock's storage of item and associative information in a composite memory representation affords the possibility to learn both contextual and order information into a composite lexical representation if the same ideas were adapted to learn from language.

A holographic account of word meaning

Convolution is basically a method of compressing the outer-product matrix of two vectors; the convolution of two vectors produces a third vector that does not resemble either argument vector, but is rather a key storing their association. When one member of the learned pair is later encountered in the environment and compared to the associative key (via correlation), the other member of the learned pair is reconstructed. Such a process is very useful because an object can be retrieved without ever storing it—it is reconstructed from an item in the environment and a stored association. Further, several

pairs of associations can be summed together in the same memory vector. Because convolution distributes over addition, a single representation can be used to represent several associative keys. Once again, when one member of a learned pair is correlated with the representation, the other member is reconstructed; if an unknown item is correlated with the representation, however, no known item can be retrieved. Such convolution–correlation memory models are often referred to as *holographic* models because they are based on the same mathematical principles used in light holography (see Plate, 2003 for a review).

A common problem with aperiodic (linear) convolution is that the associative representation is $2n - 1$ dimensions larger than the vectors representing the items themselves (where n is the dimensionality of the item vectors). For example, the convolution of two item vectors, \mathbf{x} and \mathbf{y} is:

$$x * y = z = \sum_{j=-\frac{(n-1)}{2}}^{\frac{(n-1)}{2}} x_j \cdot y_{i-j}.$$

Basically, the diagonals of the outer-product matrix are summed, producing a $2n - 1$ dimensional association. Thus, vectors representing items and their associations cannot be directly summed together because they have different dimensionality. To finesse the problem, many memory models pad the item vectors with zeros to balance dimensionality (e.g., Murdock, 1982), or simply truncate the association vector by trimming the outside elements to match the dimensionality of the item vectors (e.g., Metcalfe-Eich, 1982).

Although padding and truncation are adequate solutions for models of paired-associate learning, neither is appropriate for application to unconstrained language. Such patches still limit convolution to be used to learn pairwise associations and will miss the higher-order temporal structure that is characteristic of natural languages. To recursively bind together vectors representing all words in sentences (without expanding dimensionality) we employ *circular convolution*, a technique used extensively in image and signal processing (Gabel & Roberts, 1973; see also Plate, 1995, 2003 for examples in cognitive modeling). The circular convolution of k n -dimensional vectors always produces an n -dimensional association vector, without wasting information by truncating or expanding dimensionality by padding.¹

¹ In practice, we use a NAG fast fourier transform (FFT) version of convolution for computational efficiency. Convolution via FFT takes $O(n \log_n)$ time to compute, whereas the technique with modulo subscripts takes $O(n^2)$ time to compute. In addition, our version of convolution is non-commutative (see Plate, 1995); thus the associations formed are asymmetric (i.e., $e_1 \otimes e_2$ produces a unique vector from $e_2 \otimes e_1$).

$$x \otimes y = z = \sum_{j=0}^{n-1} x_{j \bmod n} \cdot y_{(i-j) \bmod n}.$$

Circular convolution is also referred to as *cyclic* or *wrapped* convolution because, rather than summing linearly down each diagonal of the outer-product matrix, the summing wraps around the diagonals in modulo- n steps. Hence, all elements in the matrix are summed, but dimensionality remains constant.

The BEAGLE model

We will use the circular convolution algorithm to learn associations between words in a memory model we call BEAGLE (Bound Encoding of the Aggregate Language Environment). BEAGLE constructs distributed representations for words from experience with a large-scale text corpus (text will be read and processed in one-sentence increments). The resulting representation will contain roughly the types of information inherent in both HAL and LSA, stored in a composite holographic lexicon. For example, BEAGLE will learn the types of words that share contextual information with *hammer*, and the types of words that share associative information (position relative to other words) with *hammer*.

The first time a word is encountered when reading the text corpus it is assigned a random environmental vector, e_i , which represents its physical characteristics (e.g., orthography or phonology). At this point, we are agnostic about the actual environmental structure; hence, we assume no structural similarities between words, and represent each with a different random representation. Environmental vector elements are sampled at random from a Gaussian distribution with $\mu = 0$ and $\sigma = 1/\sqrt{D}$, where D is the vector dimensionality.² Each

² The dimensionality of the vectors, D , may be set arbitrarily. Providing there are enough dimensions to adequately represent the order and context variability of words, adding dimensions does not improve or degrade the similarity structure of words in the lexicon. A version of the model trained with 1000 dimensions can produce the same pattern of similarities between words as a version trained with 10,000 dimensions. In models such as LSA, however, the model's performance can be particularly sensitive to the dimensionality selected; the model may perform better with added dimensions up to a point, but then degrades if too many dimensions are allowed (Landauer & Dumais, 1997). With BEAGLE, increasing dimensionality simply distributes the information across more values (similar to increasing the number of nodes in a connectionist network—a sufficient number is required, but additional nodes tend not to degrade performance). We recommend that a minimum of 1000 dimensions be used (1024 or 2048 are a good cache-blocking numbers), and more if one wishes to accurately retrieve word transitions from the representations.

time a word is encountered while reading the text corpus, the same environmental vector is used to represent it.

A word's memory representation, m_i , however, is adapted each sentence in which the word occurs by adding the sentence context to it. A word's context in a sentence, c , is simply the sum of the environmental representations for the other words in the sentence:³

$$\text{for } i = 1 \text{ to } N_words : c_i = \sum_{j=1}^{N_Words} e_j, \text{ where } i \neq j.$$

This new context is then added to the word's memory representation:

$$m_i = m_i + c_{i,j}.$$

A word's memory representation, thus, develops a pattern of elements that reflects its history of co-occurrence with other words in sentences. In addition, latent similarity can be formed in the lexicon between words that have never directly co-occurred in a sentence but, nonetheless, have occurred in similar contexts (around the same words) during learning. This is analogous to a latent relationship in LSA, but the relationship simply emerges from context accumulation (summing of similar random vectors) rather than SVD.

At the same time as context information is being learned for the sentence being processed so is order information, that is, information about the word's position relative to other words in the sentence. A word's order information is formed by binding it with all n -gram chunks in the sentence that include it with directional circular convolution. The position of the word being coded is represented by a constant random placeholder vector, Φ (sampled from the same element distribution as were the environmental vector elements). Each n -gram association is unique. For example, $e_1 \otimes e_2$ produces a different vector from $e_1 \otimes e_2 \otimes e_3$, (i.e., an association for a trigram is different than that for a bigram, even if the trigram contains the bigram) but both operations produce fixed-dimensional vectors so they can be directly compared and stored.

Because circular convolution is used, all n -gram associations are represented in the same fixed dimensionality and, hence, they can all be summed into a single order vector that represents the word's position relative to

³ When computing context information, function words are discarded using a stop list of 280 words the model is instructed to ignore. No stop list is used in calculation of order information (as function words play important grammatical roles). The model can in principle learn which words to ignore while it learns from experience and a continuous entropy function (see Jones & Mewhort, in press), but we use the stop list method for the simulations reported here.

all other words in the sentence. The order information, o , for a word in a sentence is thus:

$$\text{for } i = 1 \text{ to } N_words : \quad o_i = \sum_{j=1}^{p\lambda - (p^2 - p) - 1} b_{i,j},$$

where p is the position of the word in the sentence, and $b_{i,j}$ (b for “binding”) is the j th convolution chunk for the word being coded.⁴

For example, consider coding the memory representation for *excellent* in the simple sentence “dingoes make excellent pets.” The memory representation for *excellent*, $m_{\text{excellent}}$, is updated by adding the word’s context and order information from the new sentence, coded from the environmental representations for the other words:⁵

$$\begin{aligned} m_i &= m_i + c_i + o_i \\ &= m_i + \sum_{j=1}^N e_j + \sum_{j=1}^{p\lambda - (p^2 - p) - 1} b_{i,j} \\ &= m_i + \begin{bmatrix} e_{\text{dingoes}} + \\ e_{\text{make}} + \\ e_{\text{pets}} \end{bmatrix} + \begin{bmatrix} (\Phi \otimes e_{\text{pets}}) + \\ (e_{\text{make}} \otimes \Phi) + \\ (e_{\text{make}} \otimes \Phi \otimes e_{\text{pets}}) + \\ (e_{\text{dingoes}} \otimes e_{\text{make}} \otimes \Phi) + \\ (e_{\text{dingoes}} \otimes e_{\text{make}} \otimes \Phi \otimes e_{\text{pets}}) \end{bmatrix}. \end{aligned}$$

The memory representation for a word, m_i , thus, becomes a pattern of elements that reflects the word’s history of co-occurrence with, and position relative to, other words in sentences. The context information alone is an approximation to the kind of semantic structure that LSA learns, and the order information alone is similar to the type of structure learned by HAL or an SRN. BEAGLE’s learning algorithms, however, allow it to learn both types of information into a single composite representation.

Table 1 demonstrates the structure learned by the context and order equations separately when BEAGLE is trained on a text corpus. For each target word (capitalized), the eight nearest neighbors for each space are displayed (i.e., the eight words that have developed the most similar memory representations to the target). When comparing words learned by only context information, for example, *bird* is most similar to associated

words, such as *wings*, *beak*, and *nest*. In the context-only lexicon, verbs are similar to the nouns they operate upon. For example, *food* is related to *eat*, *car* is related to *drive*, and *book* is related to *read*, but *eat*, *drive*, and *read* are not highly related to one another, nor are *food*, *car*, or *book*. By contrast, when comparing words learned by the only order information, *bird* is most similar to other animals. In the order-only lexicon, words that appear in similar positions to other words in sentences develop similar structure from accumulation of common associations during learning. *Drive*, *eat* and *read* are all similar to one another, and cluster distinctly from the nouns (*car*, *food*, and *book* now being similar to one another).

The representations learned by BEAGLE are basically a blend of these two types of structure. The model contains information learned by both LSA and HAL, from very simple summation and association mechanisms and without the need for dimensional optimization. Unlike HAL, BEAGLE explicitly encodes order relations, rather than tabulating distances. Both types of information are stored together as a single composite vector pattern. Jones and Mewhort (in press) have demonstrated that BEAGLE’s composite representations more closely predict semantic relatedness in Miller’s (1995) WordNet, than does LSA. In addition, the composite representation is as good a predictor of WordNet measures as both of its component representations taken together. Hence, compressing context and order information into a single composite representation does not seem to interfere with either type of information.

Further, order sequences that were learned can be retrieved from the lexicon using inverse convolution (much in the same way Murdock (1982, 1992) retrieves items given a cue) allowing the model to perform a variety of word-transition tasks without the need for built-in transition rules. The focus of the present paper, however, will only examine the structure of the learned lexical representations, and does not require decoding of word transitions. For more information on decoding equations in BEAGLE and predicting word transitions in sentences, see Jones and Mewhort (in press).

Comparing model structure to data structure

In this section, we compare the similarity structure of representations learned by HAL, LSA, and BEAGLE to response latency data from human subjects in a range of semantic priming experiments. Of particular interest are experiments examining “purely” semantic overlap between primes and targets, associative-only prime-target relationships, and mediated prime-target relationships.

⁴ Lambda (λ) is a parameter that sets the maximum number of neighbors a word can be bound with because number of possible bindings increases symmetrically as a non-linear function of both position and sentence length towards the middle of the sentence. For the simulations reported here, λ was set to seven (after Miller & Selfridge, 1950).

⁵ The context and order vectors are each normalized to a length of one before being summed together so the two types of information are evenly weighted in the composite representation.

Table 1
Nearest neighbors to a target word in context-only and order-only lexical space

Target	Space	Nearest neighbors
Bird	Context	Wings, birds, beak, nest, fly, flying, eagle, feathers
	Order	Dog, snake, cat, child, mouse, lion, frog, balloon
Hospital	Context	Doctor, medical, patient, patients, nurse, emergency, physician, care
	Order	Library, village, clinic, zoo, barn, garage, store, house
Driving	Context	Drive, car, driver, road, truck, highway, oncoming, traffic
	Order	Making, running, getting, reading, moving, taking, riding, flying
Freud	Context	Ego, superego, personality, unconscious, freudian, psychosexual, Sigmund, theory
	Order	Lebon, Confucius, Marx, Aristotle, Muhammad, Plato, he, Gandhi
Dentist	Context	Dental, floss, gums, hygienist, plaque, cavities, teeth, brushing
	Order	Patient, teacher, reader, customer, doctor, speaker, buyer, computer
Greece	Context	Greeks, romans, civilization, ancient, ruled, rome, empire, citystate
	Order	Egypt, China, France, India, Rome, Italy, California, Brazil
Loud	Context	Heard, hear, sound, voice, noise, muffled, tone, quiet
	Order	Small, large, little, fine, slight, big, strange, strong
Bake	Context	Baked, cake, cookies, bread, flour, oven, bakers, icing
	Order	Develop, produce, create, establish, build, make, prepare, organize
English	Context	Language, speak, languages, french, spoken, dialect, england, colonies
	Order	Spanish, French, British, American, Chinese, Japanese, Indians, Russians
Sharp	Context	Teeth, cutting, cut, nose, face, edge, blade, pointed
	Order	Little, small, big, huge, broad, large, thin, solid

For the simulations reported in this paper, all three models were trained on the same text corpus, compiled by Touchstone Applied Science Associates (TASA); the TASA corpus contains approximately the quantity and quality of text that the average college-level student has experienced across his lifetime.⁶ For LSA, SVD was performed on 37,000 documents, and the 300 dimensions with the highest singular values were retained. This approximate version of LSA is available online (<http://lsa.colorado.edu>). All word comparisons reported are term to term. For BEAGLE, vector dimensionality was set a priori to 1024, and the maximum chunking parameter (λ) was set to seven (BEAGLE contained 90,000 lexical entries). In HAL, a 10-word moving window was used to implement the model, and the dimensionality of the final concatenated word vectors was reduced by retaining only the 1000 dimensions with the greatest variance (see Lund & Burgess, 1996). In addition, a standard frequency criterion was used in HAL, where words are only included in coding if they occur with a sufficient frequency. Thus, HAL has fewer lexical entries than do LSA or BEAGLE and, hence, will ultimately have missing data for some of the compari-

sons. This approximate version of HAL is available online (<http://hal.ucr.edu>).

Both LSA and BEAGLE produce dense distributed representations for words; that is, a word is represented by a complete vector of 300 or 1024 (respectively) real numbers. Thus, similarity between representations for words in either LSA or BEAGLE will be computed as the cosine of the angle between their length-normalized vectors (*vector cosine*; i.e., a normalized dot-product). Cosine is the metric for LSA provided by the web interface. HAL, on the other hand, produces sparse vector representations for words, and Euclidean distance will be used as a measure of similarity for HAL representations. Distance is also the metric provided for HAL on its web interface. Using cosine for LSA and distance for HAL allows the reader to easily validate intuitions about data presented in this paper via the online web interfaces. It is important to note, however, that cosine is a measure of similarity while distance is a measure of dissimilarity. Hence, similarity between words in LSA or BEAGLE will be reflected by a *high* cosine, whereas similarity between words in HAL will be reflected by a *low* distance metric.

Throughout this paper, we will present predictions from BEAGLE's composite lexicon, with context and order learned together for each sentence experienced.

⁶ We are grateful to Tom Landuaer for providing the TASA corpus.

We will also give predictions from BEAGLE's lexicon when trained on context information alone and when trained on order information alone. Separating the two types of learning allows comparison of the concordance between BEAGLE's components and LSA/HAL. Presenting the sources separately also demonstrates which type of information is more dominant at driving a particular priming effect in the composite representation. When we refer to BEAGLE, by default we mean the composite lexicon unless we explicitly refer to the context-only or order-only lexicons.

It is important to note that we are predicting semantic priming data directly from the structure of the lexical representations, ignoring any process explanation. That is certainly not to imply that we do not believe process plays a significant role in semantic priming. Rather, we intend to identify the structural representation model that can best explain the data directly from its representations. We can then examine process models to be applied to these structural representations, without needing to build unnecessary complexity into the process account to explain particular priming effects.

Semantic and associative priming

It has been debated whether semantic priming is the product of semantic overlap or simply association strength between the prime and target. The overall consensus (see Huchinson, 2003; Lucas, 2000; McNamara, 2005) is that there is facilitation for both prime-target pairs that share a purely semantic relationship, and for pairs that share a purely associative relationship. Generally, the facilitation for purely semantic relationships seems to be larger than that for purely associated relationships. Further, there exists an *associative boost* for prime-target pairs that are semantically similar but also have a strong association.

Exactly what is meant by a purely semantic or associative relationship between words has evolved in the past 20 years. Traditionally, semantically related words were defined as words with overlapping features or words from the same superordinate category (e.g., Battig & Montague, 1969) that were not produced as responses in association norms (e.g., Nelson, McEvoy, & Schreiber, 1998). By contrast, associated words are produced as responses to one another in free-association norms, but are not categorically similar (nor do they have overlapping semantic features). However, the semantic-or-associated distinction is more likely to be an artificial dichotomy (Huchinson, 2003; McNamara, 2005; Steyvers, 2000); there are unlikely to be any words that are purely semantically or associatively related. Nonetheless, it is possible to select pools of words that are similar from predominantly semantic or associated sources. In addition, it is now widely accepted that object-part relations (*lion-mane*) and functional relations (*broom-sweep*) are

“semantic” if they may be used to define a word (McNamara, 2005; McRae & Boisvert, 1998).

Chiarello, Burgess, Richards, and Pollok (1990)

Chiarello et al. (1990) tested prime-target pairs that had a relationship that was either semantic only (e.g., *deer-pony*), associated only (e.g., *bee-honey*), or pairs that had a relationship that was both semantic and associative (e.g., *doctor-nurse*). Semantically similar pairs were from the same superordinate category but were not related in association norms.⁷ Thus, they are expected to occur in similar contexts (and positions relative to other words), but not together in the same context. Associative pairs were related in association norms, but were not categorically similar. Thus, they are expected to occur in consistent positions relative to each other within the same contexts. The pairs that share both types of relationships are found in both the same and similar contexts, and are expected to show an associative boost relative to the semantic-only condition. BEAGLE naturally predicts an associative boost because both types of information are summed together in the model during learning. It is less intuitive, however, what predictions HAL or LSA would make for the “both” condition.

Chiarello et al. (1990) measured performance for presentations separately for each visual field; for simplicity, we will collapse across the visual fields. To make predictions from semantic space models, we prefer data from naming studies. The magnitude of effects found in lexical decision can be task specific, largely depending on the nature of the non-word foils used (Neely, 1991), and we cannot simulate non-words in a semantic space model. Hence, we will focus on Chiarello et al.'s Experiment 2 here (naming latency).

Chiarello et al. (1990) found robust facilitation in the semantic-only condition and an associative boost in the both condition. They did not, however, find a reliable priming effect in the associated-only condition. Nonetheless, they found an associated-only effect using the same stimuli in their first experiment (lexical decision), and the subtle associative effect has been demonstrated elsewhere (e.g., Perea & Rosa, 2002; Shelton & Martin, 1992; Williams, 1996), so we are sufficiently confident it exists.

Table 2 displays facilitation for each condition reported by Chiarello et al. (1990), along with predictions made by each of the models. To make model predictions, the similarity between vectors for the related prime-target pairs was subtracted from the similarity between vectors for the same target word and an unre-

⁷ However, Steyvers (2000) has noted that many of Chiarello et al.'s “semantic-only” pairs were, in fact, often associates in the Nelson et al. (1998) norms.

Table 2
Data from Chiarello et al. (1990) Experiment 2 and corresponding model predictions (Unrelated–Related in each condition)

Model	Semantic	Associated	Both
Human	+22	–3	+28
BEAGLE			
Composite	–.1673	–.0961	–.2598
Context	–.1619	–.1778	–.3005
Order	–.1567	–.0155	–.2092
LSA	–.2225	–.3400	–.4317
HAL	.3120	.1655	.4599

lated prime. The unrelated condition was simulated by pairing the same target word with a randomly selected prime word from the same set. Hence, a negative cosine difference in LSA and BEAGLE predicts facilitation: the related pairs are more similar than are the unrelated pairs. In HAL, a positive distance difference predicts facilitation: the related pairs are closer together in the space than are the unrelated pairs. All comparisons in this paper will be by items (i.e., with words as a random-effects factor), and will be compared to the equivalent items analysis in the respective experiments (although analyses by items and by subjects in the behavioral data were always in high agreement).

All three models predicted facilitation in all three conditions. HAL and BEAGLE correctly predicted a subtle associative effect, a stronger semantic effect, and the associative boost in the both condition. The magnitude of facilitation was predicted to be greater for the semantic-only condition than the associated-only condition in both BEAGLE, $t_{94} = 2.01$, $p < .05$, and HAL, $t_{63} = 2.46$, $p < .01$. LSA, on the other hand, predicted greater facilitation in the associative-only condition than it did in the semantic-only condition ($t_{94} = 2.66$, $p < .01$), which is the opposite of the empirical data demonstrating that semantic facilitation is stronger than associative facilitation for these stimuli.

An examination of the two separate components of BEAGLE demonstrates why it makes this prediction. When trained on order information only, BEAGLE resembles HAL, predicting greater facilitation in the semantic-only condition than in the associated-only condition, $t_{94} = 2.74$, $p < .01$, but it does not predict a reliable associative boost in the both condition (relative to the semantic-only condition) as HAL does, $t_{94} = 1.05$, $p \approx .29$. When trained on context information only, the pattern of cosines in BEAGLE more closely resembles that of LSA, but the effects are not reliable. Context-only BEAGLE does not predict a facilitation difference between semantic- and associated-only conditions, whereas LSA predicts greater associated facilitation than semantic facilitation. This discrepancy is most likely due to the different notions of context in

the two models: LSA considers context to be a paragraph, and BEAGLE considers context to be a sentence. When learned together in BEAGLE, the two sources of information (context and order) predict the correct pattern of facilitation across the conditions, whereas neither of them separately matches the human data.

It appears that LSA's focus on contextual co-occurrence during learning may cause the model to weight associative relationships more strongly than humans do. Further, a consideration of order information, as in HAL and BEAGLE, may be a necessary component to account for the way humans weight purely semantic relationships and to dampen purely associative relationships. To examine this possibility further, we made predictions for another experiment that attempted to separate semantic and associative prime-target relationships.

Ferrand and New (2003)

Similar to Chiarello et al. (1990), Ferrand and New (2003) examined prime-target pairs that had a relationship that was either semantic only (*whale–dolphin*) or associative only (*spider–web*), but they were more careful to select pairs that were predominantly semantically or associatively related. Further, they were more stringent on balancing orthography and phonology across conditions than the Chiarello et al. study. In calibrating their stimuli, Ferrand and New had subjects rate word pairs on how similar they were in meaning. They selected semantic-only pairs to have high semantic similarity ratings, but to be very low in associative strength based on association norms. The associative-only pairs were selected to be low in semantic similarity and high in associative strength. As a result, many of their pairs in the associated only condition were collocates. Ferrand and New used French stimuli and association norms—comparison to semantic space models trained on English text is also an interesting test of cross-language semantic stability. We used the English translations of their word pairs and computed prime-target cosines for each semantic space model. The results are presented in Table 3. For simplicity, we have collapsed over SOAs in Table 3.

The semantic-only prime-target pairs used by Ferrand and New (2003) produced greater facilitation than did the associated-only pairs. However, because they analyzed the prime-target relationships separately, it is difficult to know if this difference was statistically reliable. All models predicted facilitation in both conditions (i.e., all Unrelated–Related pairwise comparisons were significant, $p < .01$). BEAGLE and HAL both correctly predicted greater facilitation for the semantic-only pairs than for the associated-only pairs, although the difference was not significant in either model. LSA predicted greater facilitation for the associated-only pairs than for the semantic-only pairs, the opposite of the pattern of cell means in the human data and the other two models.

Table 3
Data from Ferrand and New (2003) and corresponding model predictions (Unrelated–Related in each condition)

Model	Semantic	Associated
Human	+34.67	+18.33
BEAGLE		
Composite	–.1162	–.0850
Context	–.1470	–.1659
Order	–.0554	–.0064
LSA	–.2721	–.3407
HAL	.2550	.1280

Note. We have averaged over SOAs from Ferrand and New's experiment.

However, this difference in LSA was only marginally significant, $t_{84} = 1.56$, $p \approx .10$.

Examining learning for the two components of BEAGLE separately offers insight into the relative contribution of context and order information in the composite lexicon. When trained on context information only, the predictions from BEAGLE resemble those from LSA: The associated-only pairs are slightly more similar to one another than are the semantic-only pairs, although the difference is not significant. When trained on order information only, BEAGLE predicts facilitation for only the semantically similar pairs, $t_{82} = 2.11$, $p < .05$, but not for the associated pairs, $t_{82} = 0.21$, $p \approx .83$, a pattern more consistent with the predictions from HAL.

Similar to the simulations using the Chiarello et al. (1990) data, it appears that LSA may be overweighting associative relationships, and underweighting semantic relationships, relative to the human data. An examination of BEAGLE's two components suggests that this sensitivity to associative relationships is due to LSA's focus on contextual co-occurrences during learning. Further, the order information in BEAGLE is more similar to HAL and the human data, and appears to be driving BEAGLE's correct pattern for semantic and associative relationships in the composite lexicon. To explore this possibility further, we examined priming experiments that used various types of associative relationships, and the ability of the models to make accurate predictions of the observed human data in those experiments.

Williams (1996): Phrasal associates

Williams (1996) examined prime-target pairs that had various types of relationships: semantically similar pairs with low association strength (e.g., *road–path*), category coordinates (e.g., *cup–saucer*), collocates that were part of a phrase (e.g., *needle–thread*), and associates that were not phrasal collocates (e.g., *cow–milk*). Both collocates and associates had high strength in association norms, but the collocates also tended to appear in familiar con-

junctive phrases whereas the associates did not (as determined by Williams' subjective ratings). Huchinson (2003) has noted, however, that many of Williams' associated pairs were also of the same superordinate semantic category. We will focus on pronunciation latencies in Williams' intact targets condition only.

Table 4 shows facilitation for each condition in Williams' (1996) study (in ms), and corresponding predictions from representation similarity in each of the three semantic space models. Williams found significant facilitation priming in all four conditions, however, the collocates displayed greater facilitation than was observed in the other conditions. Most importantly for Williams' stimuli was the finding that collocates produced significantly greater facilitation than did the associates that did not have a collocational relationship. Both HAL and BEAGLE predict the general facilitation differences between conditions found by Williams. Specifically, BEAGLE predicted Williams' finding of greater facilitation for the collocate pairs than the associated-only pairs, $t_{30} = 2.18$, $p < .05$. HAL predicted the general trend, however, collocates were only marginally more similar to each other than were associated-only pairs, $t_{30} = 1.67$, $p \approx .10$. LSA predicted strong facilitation for both collocates and associated-only pairs, with no significant differences between the two groups, $t_{30} = 0.32$, $p \approx .75$.

The context learning in BEAGLE was consistent with predictions from LSA, predicting the same facilitation for both collocates and non-collocational associates, $t_{30} = 0.32$, $p \approx .75$. The order information in BEAGLE was correctly driving the difference between collocates and associates, with greater predicted facilitation for collocates than associates, $t_{30} = 2.34$, $p < .05$.

LSA learns information about statistical co-occurrences in contexts across a text corpus. It does not, however, pay attention to word order information within those contexts. Thus, LSA predicts equivalent facilitation to associates that appear in the same context (*bee–honey*) and associates that are related by virtue of both being members of a common collocational phrase (e.g., *surf–web*, *pay–attention*). However, humans clearly pay attention to this word-order structure within contexts during learning and, hence, collocational relationships appear stronger in human semantic memory than are associative relationships that are not collocational. Both HAL and BEAGLE pay attention to word-order information when learning a text corpus (albeit, in very different ways), which allows stronger collocational similarities than simple associative relationships in their respective lexicons.

Moss, Ostrin, Tyler, and Marslen-Wilson (1995): Functional relations

Moss et al. (1995) demonstrated that priming also occurs for functionally related but non-associated

Table 4
Data from Williams (1996) and corresponding model predictions (Unrelated–Related in each condition)

Model	Semantic	Category coordinate	Collocates	Associates
Human	+1	+4	+18	+3
BEAGLE				
Composite	-.1312	-.1485	-.2771	-.1778
Context	-.2129	-.2308	-.2570	-.2391
Order	-.0861	-.0682	-.3014	-.1091
LSA	-.3382	-.3056	-.4519	-.4388
HAL	.2631	.2227	.4284	.2360

words. In their experiment (Experiment 1: auditory priming), they independently manipulated type of prime-target relationship (categorical, functional) and the normative association between prime and target (associated or non-associated). Within the category coordinates condition, they used pairs that were from natural categories (associated: *thunder–lightning*, non-associated: *aunt–nephew*) or from artificial categories (associated: *bat–ball*, non-associated: *dagger–spear*). Within the functional relations condition, they used pairs that were either script relations (associated: *beach–sand*, non-associated: *party–music*) or instrumental relations (associated: *bow–arrow*, non-associated: *broom–floor*). Moss et al. found significant facilitation for both categorically related prime-target pairs and for prime-target pairs that were functionally related. In addition, they found an associative boost whereby prime-target pairs with a normative association produced greater facilitation than did non-associative prime-target pairs in both categorically- and functionally related conditions.

Table 5 shows facilitation (in milliseconds) for each specific condition in the Moss et al. (1995) study, along with corresponding predictions for the same word pairs from each of the three semantic space models. As in our previous simulations, we subtracted the related prime-target similarities from the similarities between the same target and a randomly re-paired prime from the same condition on a pairwise basis.

Generally, all three models predicted facilitation in both categorically-related and functionally related conditions. In BEAGLE, the related prime-target pairs had significantly higher cosines than did the unrelated pairs, both for categorical relationships ($M_{\text{related}} = 0.4953$, $M_{\text{unrelated}} = 0.2813$, $F(1,55) = 65.36$, $p < .001$) and for functional relationships ($M_{\text{related}} = 0.3747$, $M_{\text{unrelated}} = 0.2948$, $F(1,54) = 18.78$, $p < .001$). LSA produced the same pattern, with higher cosines between related pairs than unrelated pairs both for categorical relationships ($M_{\text{related}} = 0.4487$, $M_{\text{unrelated}} = 0.0791$, $F(1,55) = 152.69$, $p < .001$) and functional relationships ($M_{\text{related}} = .3582$, $M_{\text{unrelated}} = 0.0701$, $F(1,55) = 103.24$, $p < .001$). HAL also predicted facilitation in

both conditions, with related pairs being significantly closer to each other than unrelated pairs, both for categorical relationships ($M_{\text{related}} = 1.27$, $M_{\text{unrelated}} = 1.62$, $F(1,37) = 68.67$, $p < .001$) and functional relationships ($M_{\text{related}} = 1.53$, $M_{\text{unrelated}} = 1.68$, $F(1,34) = 21.11$, $p < .001$).

All three models correctly predict facilitation using both the categorically- and the functionally related stimuli from Moss et al. (1995). However, Moss et al. found a null effect of relationship (categorical/functional) on facilitation, but all three models produce stronger similarity between categorically related pairs than between functionally related pairs. It is not directly obvious why categorical pairs are more similar than functional relations in all models; for now, we will leave this as an anomaly and proceed to examine the associative boost separately for category coordinates and functional relations.

The most important finding from the Moss et al. (1995) study was the *associative boost*: for both category coordinates and functional relations, prime-target pairs that also had a normative association produced greater facilitation than did non-associative pairs. The predictions from BEAGLE were the most similar to the human data. For category coordinates in BEAGLE, predicted facilitation was significantly greater for associated prime-target pairs ($M = -0.2410$) than for non-associated pairs ($M = -0.1869$), $F(1,54) = 4.13$, $p < .05$. For functional relations in BEAGLE, facilitation predicted for associated pairs ($M = -0.1002$) was greater than that predicted for non-associated pairs ($M = -0.0587$), but this difference was only marginally significant, $F(1,53) = 2.31$, n.s. By contrast, predicted facilitation for category coordinates in LSA did not differ between associated ($M = -0.3886$) and non-associated ($M = -0.3507$) pairs, $F(1,54) = 0.39$, n.s., but for functional relations, LSA correctly predicted greater facilitation for associated pairs ($M = -0.3618$) over non-associated pairs ($M = -0.2413$), $F(1,54) = 7.58$, $p < .01$. In HAL, there were no differences in facilitation between associated and non-associated prime-target pairs for either category coordinates or functional relations (both $F < 1$).

BEAGLE correctly predicts an associative boost in the category coordinates condition. It also predicts a

Table 5
Data from Moss et al. (1995) and corresponding model predictions (Unrelated–Related in each condition)

Model	Category coordinates				Functional relations			
	Associated		Non-associated		Associated		Non-associated	
	Natural	Artificial	Natural	Artificial	Script	Instrumental	Script	Instrumental
Human	+122	+95	+88	+88	+111	+105	+76	+54
BEAGLE								
Composite	–.3427	–.1392	–.1876	–.1762	–.1080	–.0925	–.0705	–.0478
Context	–.3815	–.1981	–.1958	–.2122	–.2101	–.1574	–.1865	–.0793
Order	–.3222	–.0887	–.1818	–.1206	.0030	–.0068	.0184	–.0209
LSA	–.4621	–.3450	–.3329	–.3696	–.3764	–.3471	–.2343	–.1943
HAL	.4614	.1767	.3966	.3375	.2136	.0945	.2632	.0098

pattern consistent with the associative boost in the functional relations condition, although this difference was only marginally significant. LSA predicts a strong associative boost for functional relations that frequently appear in the same context, but cannot detect any benefit of normative association for category coordinates as the human subjects did. HAL was expected to produce the opposite prediction from LSA: an associative boost in the category coordinates condition but not in the functional relations condition. However, HAL did not predict an associative boost in either condition. In addition, all three models correctly favor script relations over instrumental relations and natural over artificial category relations, whether associated or not.

Summary of “pure” semantic and associative priming

All three semantic space models are capable of predicting facilitation due to semantic overlap between primes and targets. Generally, LSA overestimates the strength of purely associative priming compared to purely semantic priming. Certain types of associative relationships, such as collocational phrases, can produce greater facilitation than do purely semantic relationships. However, LSA ignores information about word order during learning that would reveal differences between collocates and regular associates, and this is clearly a source of information that is salient to humans, as is indicated by the priming data.

All three models predict facilitation both to the category coordinates and to the functional relations from Moss et al. (1995); however, the important associative boost was more elusive. HAL did not distinguish between associated and non-associated for either categorically related or functionally related pairs. LSA was sensitive to the associative boost for functional relations only. The composite representation in BEAGLE produced a pattern that was most consistent with the human data: an associative boost for both categorical relations and functional relations, although the functional associative boost was considerably smaller than

that predicted by LSA. At this point, we may conclude that a model that pays attention to both context and word order while learning stands the greatest chance of matching the trends found in the aggregate human data.

Mediated priming

In mediated priming, the relationship between the prime and target (e.g., *lion-stripes*) is through a mediated concept (e.g., *tiger*). Mediated priming often produces only a subtle effect on response latency, an effect that is more pronounced in naming than lexical decision (Balota & Lorch, 1986; McNamara & Altarriba, 1988). However, even when it is not observed in latency data, mediated priming is still observed in neurocognitive measures of brain function, and is associated with the N400 component in evoked brain potential recordings (Chwilla, Kolk, & Mulder, 2000).

Mediated priming has been taken as evidence in favor of semantic network models and against feature lists. In a network representation, activation of a node will spread activation to related nodes. Thus, when the *lion* node is activated, the activation spreads to *tiger* because it is closely connected to *lion*, and then to *stripes* because it is closely connected to *tiger*. The effect is difficult to explain with feature-list representations because *lion* and *stripes* should have few directly overlapping features (in fact, *stripes* should be a feature slot, not a distinct representation).

Alternatively, McKoon and Ratcliff (1988, 1992) have suggested a compound cue explanation for the effect: mediated priming may not be “mediated” per se, but rather may reflect a direct associative relationship between the prime and target from lexical co-occurrence. Semantic space models would predict mediated priming for a similar reason to McKoon and Ratcliff’s notion. The information that becomes shared between *lion* and *stripes* is learned from their statistical co-occurrence in language, which is mediated through their respective relationships to *tiger*. In the learned representations,

however, the similarity between *lion* and *stripes* is direct, and exists even without a lexical entry for *tiger*. Thus, similar to McKoon and Ratcliff, the prime-target similarity is direct rather than mediated. Unlike McKoon and Ratcliff, however, semantic space models do not require the two words to have been simultaneously present in short-term memory during encoding—indirect information can become shared between two words that have never directly co-occurred in the same context during learning.

Livesay and Burgess (1998a) have tested mediated prime-target pairs in HAL and found that the model does not predict facilitation; in fact, mediated pairs are actually less similar to each other than are unrelated pairs in HAL (Livesay & Burgess, 1998b). The finding has been taken to suggest that mediated priming is not due to weak direct relationships learned from co-occurrences in language. However, the null finding may reflect the specific type of co-occurrence information that HAL learns. For example, Chwilla and Kolk (2002) have reported evidence of mediated similarity using LSA, and this finding has been corroborated by Huchinson (2003). In this section, we will test similarity predictions from all three semantic space models in their accordance to human data from mediated priming experiments. The pattern of results may help resolve the type of information that a model must consider during learning to account for mediated priming, or if a semantic network representation is the only feasible architecture to explain the effect.

Balota and Lorch (1986): Mediated priming in word naming

Balota and Lorch (1986) examined prime-target pairs that had either a direct relationship (e.g., *tiger-stripes*) or a relationship that was mediated through a concept related to both (e.g., *lion-stripes*). They used the same targets in all conditions, and the unrelated baseline condition used the same target paired with an unrelated prime from the same set (e.g., *breeze-stripes*). Representation structure for the same prime-target pairs was computed for each semantic space model; the results are presented in Table 6.

Both BEAGLE and LSA predicted the mediated priming effect found by Balota and Lorch; that is, the mediated pairs were significantly more similar than were the unrelated pairs in both BEAGLE $t_{47} = 2.35, p < .05$, and LSA, $t_{47} = 2.66, p < .05$. By contrast, HAL predicted a null effect of mediated priming, $t_{34} = 0.62, p \approx .54$, replicating the reports of Livesay and Burgess (1998a, 1998b).

Examining BEAGLE's two learning components separately demonstrates that it is the context information (similar to LSA) that is driving the effect of mediated similarity in the composite representation, not the

Table 6

Data from Balota and Lorch (1986) and corresponding model predictions (data are RTs or prime-target cosines/distances for each condition)

Model	Related	Mediated	Unrelated
Human	549	558	575
BEAGLE			
Composite	.3647	.3196	.2653
Context	.3242	.2532	.1974
Order	.4053	.3981	.3338
LSA	.3202	.1998	.1338
HAL	1.49	1.58	1.60

order information (more similar to HAL). Mediated prime-target pairs were more similar than unrelated pairs in BEAGLE when trained on context information alone, $t_{47} = 2.75, p < .01$. By contrast, there was a null difference between mediated and unrelated pairs in BEAGLE when trained on order information alone, $t_{47} = 1.64, n.s.$ Hence, it seems that a focus on indirect contextual learning, as in LSA and BEAGLE's context information, is necessary to create representations that have the structure required for mediated similarity.

McNamara and Altarriba (1988): Mediated priming in lexical decision

McNamara and Altarriba (1988) further examined mediated priming in a lexical decision task using new stimulus pairs that were based on those used by Balota and Lorch (1986), but the mediator became the target, and the prime became the mediator (e.g., *mane-tiger*, mediated by their relationship to *lion*). Once again, we subtracted the similarity between mediated prime-target pairs from an unrelated baseline using the same target word re-paired with an unrelated prime (e.g., *pupil-tiger*).

McNamara and Altarriba (1988) found significantly faster reaction times to mediated pairs (878 ms) than unrelated pairs (915 ms). In BEAGLE, the mediated pairs were more similar ($M = 0.31$) than were the unrelated pairs ($M = 0.24$), $t_{39} = 3.43, p < .001$, as they were in LSA as well ($M_{\text{mediated}} = 0.25$, $M_{\text{unrelated}} = 0.08$), $t_{39} = 6.67, p < .001$. HAL predicted the correct direction of similarity, with mediated pairs being slightly closer ($M = 1.46$) in space than the unrelated pairs ($M = 1.58$), however, the difference was not reliable. Once again, the mediated effect in BEAGLE was produced by its context information, not its order information. When trained on context information alone, the mediated prime-target pairs ($M = 0.2496$) were significantly more similar than were the unrelated pairs ($M = 0.1257$), $t_{39} = 5.38, p < .001$. When trained on order information alone, however, there was a null dif-

ference between mediated pairs ($M = 0.3843$) and unrelated pairs ($M = 0.3101$), $t_{39} = 1.80$, n.s.

McNamara (1992): Long-distance mediated priming

McNamara (1992) tested mediated priming where the relationship between the prime and target was via a chain of two mediated concepts. For example, given the prime-target pair *lifeguard-box*, the mediating chain would be *lifeguard-beach-sand-box*. McNamara has argued that facilitation in such pairs more likely reflects spreading activation and argues against models based on statistical co-occurrence since the prime and target are unlikely to have any direct semantic or associative relationship.

McNamara (1992) found facilitation with long-distance mediated pairs using a lexical decision task, that is, targets were processed more efficiently when preceded by a mediated word (597 ms) than by an unrelated word (607 ms); the 10-ms effect was subtle, but reliable. In BEAGLE, the mediated pairs were significantly more similar ($M = 0.22$) than were the unrelated pairs ($M = 0.16$), $t_{39} = 2.14$, $p < .05$, as they were in LSA ($M_{\text{mediated}} = 0.15$, $M_{\text{unrelated}} = 0.09$), $t_{39} = 3.09$, $p < .01$. While both BEAGLE and LSA predicted the subtle effect of long-distance mediated priming, there were no similarity differences between mediated pairs ($d = 1.71$) and unrelated pairs ($d = 1.68$) in HAL.

Contrary to McNamara's (1992) suggestion, long-distance mediated priming can be accounted for in a model that simply learns statistical co-occurrences of words. The indirect relationships formed from contextual co-occurrences, as in LSA and BEAGLE's context learning, are essential to build representations with this mediated structure. As in previous simulations of mediated priming, BEAGLE's context information predicted the effect, with greater similarity between mediated pairs ($M = 0.1472$) than unrelated pairs, ($M = 0.0925$), $t_{39} = 3.00$, $p < .01$. However, BEAGLE's order information was similar to HAL, with a null similarity difference between mediated pairs ($M = 0.2755$) and unrelated pairs ($M = 0.2315$), $t_{39} = 1.17$, n.s.

de Groot (1983)

As an additional comparison of the three semantic space models, we made predictions for de Groot's (1983) one-step mediated stimuli. de Groot's stimuli were presented in Dutch, but we used the English translations to make model predictions (see her Appendix A). Using a naming task, de Groot found significant priming for both her related and mediated conditions. de Groot's results and model predictions using the English translation of her stimuli are presented in Table 7.

All three models predicted facilitation in the related condition (i.e., all U-R comparisons were significant).

Table 7

Data from de Groot (1983) and corresponding model predictions (data are RTs or prime-target cosines/distances for each condition)

Model	Related	Mediated	Unrelated
Human	494	518	537
BEAGLE			
Composite	.4590	.3464	.2892
Context	.4202	.2459	.1613
Order	.5075	.4286	.4039
LSA	.4270	.1913	.0703
HAL	1.34	1.58	1.65

For the mediated condition, the mediated prime-target pairs were significantly more similar to each other than were the unrelated pairs in BEAGLE, $t_{29} = 2.08$, $p < .05$, and LSA, $t_{29} = 3.74$, $p < .001$. HAL also predicted greater similarity between the mediated pairs than the unrelated pairs, but the difference was only marginally significant, $t_{22} = 1.83$, $p \approx .08$. Once again, the mediated effect in BEAGLE was driven by context learning. Pure context information predicted facilitation to the mediated pairs, $t_{28} = 2.81$, $p < .01$, but pure order information predicted a null effect, as in HAL, $t_{29} = 0.50$, n.s.

McKoon and Ratcliff (1992): Compound cues

McKoon and Ratcliff (1992) have argued that the phenomenon of mediated priming is not actually "mediated" at all, but reflects direct familiarity between the prime and target as a compound from their simultaneous presence in short-term memory. In their 1992 study, McKoon and Ratcliff manipulated direct familiarity between prime and target independent of associative links. To select their stimuli, McKoon and Ratcliff moved a six-word window over the AP newswire corpus and recorded the frequency of co-occurrence within the window for all word pairs. Word pairs that co-occurred with a greater than chance frequency were selected. In their stimulus set, McKoon and Ratcliff included prime-target pairs that had a semantic relation (e.g., *child-baby*), pairs that had a high probability of co-occurrence (e.g., *hospital-baby*), and words that had a low, but greater than chance, probability of co-occurrence (e.g., *room-baby*). The same target word was used in all conditions. Table 8 displays McKoon and Ratcliff's results (using a lexical decision task) with similarity predictions from each of the semantic space models using the same pairs.

McKoon and Ratcliff (1992) found faster reaction times in both the semantically related and high co-occurrence conditions when compared to the unrelated condition. The low co-occurrence condition, however, was not statistically different from the unrelated condition. All three semantic space models predicted the facilitation

Table 8

Data from McKoon and Ratcliff (1992) and corresponding model predictions (data are RTs or prime-target cosines/distances for each condition)

Model	Semantic	High <i>t</i> -value	Low <i>t</i> -value	Unrelated
Human	500	528	532	549
BEAGLE				
Composite	.5549	.4131	.3518	.2919
Context	.5024	.3366	.2836	.1701
Order	.6161	.4820	.4146	.4055
LSA	.4722	.2230	.2007	.0522
HAL	1.234	1.478	1.528	1.636

observed in the data for the semantically related and high co-occurrence conditions (i.e., all U-R comparisons were significant, $p < .01$).

For the low co-occurrence condition, BEAGLE predicts greater similarity for the low pairs than the unrelated pairs, $t_{39} = 2.21$, $p < .05$. HAL predicted only a marginal similarity difference between low co-occurrence pairs and unrelated pairs, $t_{29} = 1.79$, $p \approx .08$. LSA, however, predicted a strong priming effect in the low co-occurrence condition, $t_{39} = 6.61$, $p < .001$. The high and low co-occurrence pairs were specifically selected to be statistically different (from their probability of co-occurrence in text), however, only BEAGLE predicts greater similarity between high co-occurrence pairs than between low co-occurrence pairs, $t_{39} = 2.45$, $p < .05$; both HAL and LSA predict a null difference.

Summary of mediated priming

BEAGLE and LSA are able to accurately simulate the empirical mediated priming trends directly from prime-target representation similarity, but HAL cannot. While word-order information (considered in BEAGLE and HAL) is necessary to simulate semantic and associative priming effects, the indirect co-occurrence information (considered by BEAGLE and LSA) is clearly important to account for human data in mediated priming tasks. BEAGLE's composite representation is a blend of both types of information, allowing it to predict semantic and associative facilitation, functional priming, the associative boost, and mediated priming, all from the same representation.

General discussion

Semantic space models are particularly appealing because they learn representations for words automatically from statistical characteristics of language. The approach solves the "hand coding" problem inherent

in models of semantic representation such as feature lists and semantic networks,⁸ and leaves much of the representation complexity in environmental redundancy rather than artificially hardwiring complexity into the model's representation. However, the ability of semantic space models to account for trends in human data has not been widely studied.

We found that three models that all fit into the class of "co-occurrence" or "semantic space" models can still make subtly, and sometimes wildly, different predictions for semantic priming tasks from the structure of their learned representations, even though all three models were trained on the same text corpus. The pattern of similarities and differences between co-occurrence models raises a particularly important question: what is co-occurrence and how can it be best modeled? BEAGLE differentiates between *contextual* similarity and *order* similarity (as Murdock has done with associative learning), and incorporates both types of information in a composite representation, allowing it to best match the trends found in human priming data.

LSA focuses on contextual co-occurrences during learning. However, the strongest relationships in a final LSA matrix are often between words that have never directly co-occurred in the same context as many as 70% of a word's closest neighbors in the LSA space never directly occurred with it in context. (Landauer & Dumais, 1997). Nonetheless, LSA tends to develop stronger relationships between associates than it does between common members of a semantic category, whereas humans generally weight semantic relationships more strongly than associative relationships in priming tasks. For example, in LSA, *car* is more similar to *drive* than it is to *boat* or *truck*, and *read* is more similar to *book* than it is to *write* or *think*.

By contrast, HAL develops stronger similarity for words that have appeared similar distances from other words during learning, even if they have not directly co-occurred within the moving window. That is, positional similarity is more important to HAL and, hence, the model tends to develop greater similarity between semantic and lexical class relations than it does for associative relationships. For example, in HAL *car* is more similar to *boat* or *truck* than it is to *drive*, and *read* is more similar to *write* or *think* than it is to *book*.

BEAGLE considers both sources of information (context and word order) during learning and its representation is a blend of the two types of information. The composite representation is better able to account for

⁸ However, Steyvers and Tenenbaum (2005) have examined semantic networks that form new connections from experience based on social network theory.

the aggregate priming data from a blend of both context and order information.

This paper has demonstrated that semantic space models can indeed account for data from mediated priming experiments, but the contextual co-occurrence information that BEAGLE and LSA take into account while learning is necessary for mediated relationships. The phenomenon of mediated priming emerges from latent (indirect) relationships in the lexicon. HAL did not predict mediated priming in any of the experiments examined, whereas BEAGLE and LSA did.

Semantic space models and compound cue models (McKoon & Ratcliff, 1988, 1992) agree that spreading activation is unnecessary to explain mediated priming: the relationship between the prime and target reflects directly shared associative or semantic similarity. However, while compound cue models require the two words to have appeared together in short-term memory during learning (hence, in the same context), semantic space models do not. In both BEAGLE and LSA, words may develop latent relationships in the lexicon without necessarily occurring together in contexts. *Lion* and *stripes* may become similar to one another because they appear around the same words, such as *zebra*, *tiger*, *Africa*, *claws*, etc. They occur in similar contexts in language because of the similarities between lions and tigers in the environment, but the relationship between *lion* and *stripes* in the lexicon is direct and exists even without a lexical entry for *tiger*. *Lion* and *stripes* develop similar representations in BEAGLE and LSA for the same reason that *lion* and *cougar* would: they do not directly co-occur, but rather they occur in contexts with similar words. Given its focus on distance information, HAL would see similarity between *lion* and *tiger* or *cougar*, but less similarity between *lion* and *stripes* or even *tiger* and *stripes*. It is the contextual information in BEAGLE (the LSA-like half of the information it considers) that allows it to account for mediated priming.

BEAGLE was designed as a means of incorporating word-order information into a semantic space model such as LSA, using very simply associative mechanisms that are used to form other memory representations, and have been effective in accounting for judgments of frequency and recency (Murdock & Smith, 2001), recognition and serial recall (Mewhort & Popham, 1991; Murdock, 1992), and free recall (Franklin & Mewhort, 2002). Further, these same mechanisms allow the model to retrieve word-transition information stored into the model's lexicon (see Jones & Mewhort, *in press*).

The various types of priming examined in this paper make it clear that human semantic representation includes information about both word context and word order. BEAGLE learns both types of information into a composite representation for a word's meaning, and only this composite representation contains the neces-

sary structure to account for the broad range of priming data reported here, from differences in pure semantic and associative priming, to phrasal associates, and mediated priming. Models that focus on only one type of information will account for some priming tasks, but will fail to account for them all. Differences across types of priming provide important clues to the structure of semantic memory, and the types of information humans use when leaning that structure.

The specific learning mechanisms in the three semantic space models are not the only differences between them. Traditionally, LSA had been trained on an encyclopedia corpus (e.g., Landauer & Dumais, 1997), and HAL had been trained on a large body of usenet text (e.g., Lund & Burgess, 1996). Thus, variations in model output could be attributed to differences in the learning mechanisms, the representations, or simply differences in the input corpora. This is the first study to compare semantic space models trained on a common text corpus.

Still, the models differ broadly on what they consider a *context* to be. In LSA, a context is specifically a paragraph or article. In BEAGLE, the sentence is the context unit because order information must be computed on discrete sentences (a complete syntactic structure). While BEAGLE and LSA analyze meaningful units of context, HAL simply moves an *n*-word window across a textbase, recording local distances from within the window at each one-word increment. Given the differences in contexts between the models, it is remarkable how similarly they predict many tasks. When BEAGLE is trained using context information only (without the convolution order algorithm) on the same text corpus as LSA, the pattern of nearest neighbors is strikingly similar between the two models, considering they are using different ideas of a word's context (sentences vs. paragraphs).

In addition, BEAGLE is more stochastic than are HAL and LSA, and the final memory representations learned by BEAGLE may be particularly sensitive to initial conditions. While HAL and LSA produce the same representation for words in a textbase on multiple training epochs, BEAGLE's representations depend on the initial random environmental representations. Although two training runs will produce different representations for a word, the similarity of the word's representation to the representations of other words is remarkably consistent on different training runs, provided that the initial environmental vectors exhibit random structure. Building arbitrary similarity between initial environmental vectors can have a chaotic effect on the final memory representations produced by BEAGLE (of course, building similarity structure *consistent* with orthographic similarity structure should, in theory, produce a better memory representation).

Many models of priming focus on a description of the *process* of priming while avoiding a description of

the specific structure this process operates upon, and how the structure is learned and represented. Co-occurrence models excel at creating the structural memory representations that are needed as input to process accounts. In addition, we have demonstrated that much of the complexity needed to produce certain effects (such as mediated priming) is available in the structural representations learned by semantic space models, and need not be redundantly build into the process mechanism.

However, semantic space models certainly require an appropriate process mechanism to properly account for the real-time temporal dynamics of priming. Although spreading activation is a feasible process mechanism to operate on semantic space representations, they do not require spreading activation to explain mediated priming. Others have been successful at simulating a broad range of temporal priming effects using distributed attractor networks trained on feature list representations (e.g., Cree, McRae, & McNorgan, 1999; Plaut, 1995; Plaut & Booth, 2000). Specifically, Plaut's model represents semantic and associative similarity between words from distinct stores. Semantic relatedness between words is a product of feature overlap, and associative similarity as the likelihood that one word follows another during training. In principle, Plaut's model could account for mediated priming with its associative mechanism, but this has not been demonstrated. BEAGLE's vectors could simply be used as word representations in an attractor network such as Plaut's, allowing for word-specific predictions. The resulting model would have a realistic representation of words from text experience, and could predict semantic and associative priming, temporal effects of SOA, and mediated priming, without needing multiple representations for semantic and associative information.

Finally, we have claimed that semantic space models learn their lexical representations from statistical redundancies in the environment; however, the models have only ever experienced text. Although text is a convenient input format on which to train statistical learning models it is, at best, only an approximation of the complete environmental structure humans learn from. Many effects in semantic priming must be dependent on perceptual learning (e.g., Solomon & Barsalou, 2001), which is not represented in semantic space models trained only on text. Learning about words from sensorimotor experience is certainly not inconsistent with the mechanisms in semantic space models. Sensorimotor experience is still statistical information, but the models require appropriate front-end representations to take advantage of this additional experience, and work with small-scale perceptually grounded models has suggested that learning sensorimotor information before textual information can help ground meaning, leading to better word representations (Howell, Jankowicz, & Becker, 2005).

References

- Audet, C., & Burgess, C. (1998). Distinguishing between manner of motion and inherently directed motion verbs using a high-dimensional memory space and semantic judgments. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 66–71.
- Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 336–345.
- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 80, 1–46.
- Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In Dietrich & Markman, (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 117–156). NJ: Lawrence Erlbaum Associates, Publishers.
- Chiarello, C., Burgess, C., Richards, L., & Pollok, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't...sometimes, some places. *Brain & Language*, 38, 75–104.
- Chwilla, D. J., & Kolk, H. J. (2002). Three-step priming in lexical decision. *Memory & Cognition*, 30, 217–225.
- Chwilla, D. J., Kolk, H. J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, 42, 314–341.
- 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. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York, NY: Academic Press.
- Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, 23, 371–414.
- 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.
- de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior*, 22, 417–436.
- Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, 29, 145–193.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.
- Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. In P. Bonin (Ed.), *Mental lexicon: Some words to talk about words* (pp. 25–43). Hauppauge, NY: Nova Science Publishers.
- Franklin, D. R. J., & Mewhort, D. J. K. (2002). An analysis of immediate memory: The free-recall task. In N. J. Dimpol-

- oulos & K. F. Li (Eds.), *High performance computing systems and applications*. New York: Kluwer.
- Gabel, R. A., & Roberts, R. A. (1973). *Signals and linear systems*. New York: Wiley.
- Griffiths, T. L., Steyvers, M., Blei, D. M., & Tenenbaum, J. B. (2005). Integrating topics and syntax. *Advances in Neural Information Processing Systems*, 18, 2–5.
- Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, 53, 258–276.
- Huchinson, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10, 785–813.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220–264.
- Jones, M. N., & Mewhort, D. J. K. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, in press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. NY: Cambridge University Press.
- Laham, R. D. (2000). Automated content assessment of text using latent semantic analysis to simulate human cognition. Unpublished doctoral dissertation, University of Colorado at Boulder.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–240.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society* (pp. 412–417). Mahwah, NJ: Erlbaum.
- Livesay, K., & Burgess, C. (1998a). Mediated priming does not rely on weak semantic relatedness or local co-occurrence. In *Proceedings of the Cognitive Science Society*.
- Livesay, K., & Burgess, C. (1998b). Mediated priming in high-dimensional semantic space: No effect of direct semantic relationships or co-occurrence. *Brain and Cognition*, 37, 102–105.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, 7, 618–630.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, 203–208.
- McKoon, G., & Ratcliff, R. (1988). Assessing the occurrence of elaborative inference with recognition: Compatibility checking of compound cue theory. *Journal of Memory and Language*, 28, 547–563.
- McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1155–1172.
- McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1173–1190.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Psychology Press.
- McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language*, 27, 545–559.
- 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. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130.
- Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review*, 89, 627–661.
- Mewhort, D. J. K., & Popham, D. (1991). Serial recall of tachistoscopic letter strings. In W. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays in honor of Bennet B. Murdock* (pp. 425–443). Hillsdale, NJ: Erlbaum.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Miller, G. A. (1995). WordNet: An online lexical database. *Special Issue of International Journal of Lexicography*, 3.
- Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, 63, 176–185.
- Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 863–883.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Murdock, B. B. (1992). Item and associative information in a distributed memory model. *Journal of Mathematical Psychology*, 36, 68–98.
- Murdock, B. B. (1993). Derivations for the chunking model. *Journal of Mathematical Psychology*, 37, 421–445.
- Murdock, B. B., & Smith, D. G. (2001). Judgments of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology*, 45, 564–602.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. (1998). University of South Florida word association, rhyme and word fragment norms.
- 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.
- Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, 66, 180–194.

- Perfetti, C. A. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, 25, 363–377.
- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6, 623–641.
- Plate, T. A. (2003). Holographic Reduced Representations. CSLI Lecture Notes Number 150, CSLI Publications, Stanford, CA, ISBN 1575864304.
- Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.
- Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single mechanism account of lexical processing. *Psychological Review*, 107, 786–823.
- Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–233.
- Salton, G. (1973). Recent studies in automatic text analysis and document retrieval. *Journal of the Association of Computing Machinery*, 20, 258–278.
- Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the Association of Computing Machinery*, 18, 613–620.
- Servan-Schreiber, D. A., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning*, 7, 161–194.
- Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1191–1210.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214–241.
- Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology*, 43, 129–169.
- Steyvers, M. (2000). Modeling semantic and orthographic similarity effects on memory for individual words. Unpublished doctoral dissertation, Indiana University.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41–78.
- Wiemer-Hastings, P. (2000). Adding syntactic information to LSA. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Wiemer-Hastings, P. (2001). Rules for syntax, vectors for semantics. In *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Williams, J. N. (1996). Is automatic priming semantic? *European Journal of Cognitive Psychology*, 8, 113–161.