Semantic Spaces based on Free Association that Predict Memory Performance

Mark Steyvers Stanford University
Richard M. Shiffrin Indiana University
Douglas L. Nelson University of South Florida

Submitted to JEP General

Abstract

Many memory models represent aspects of words such as meaning by vectors of feature values, such that words with similar meanings are placed in similar regions of the semantic space whose dimensions are defined by the vector positions. Methods for constructing such spaces include those based on scaling similarity ratings for pairs of words, and those based on the analysis of co-occurrence statistics of words in contexts (Landauer & Dumais, 1997). We utilized a Word Association Space (WAS), based on a scaling of a large data base of free word associations: Words with similar associative structures were placed in similar regions of the high dimensional semantic space. In comparison to LSA and other measures based on associative strength, we showed that the similarity structure in WAS is well suited to predict similarity ratings in recognition memory, percentage correct responses in cued recall and intrusion rates in free recall. We suggest that the WAS approach is a useful and important new tool in the workshop of theorists studying semantic effects in episodic memory.

An increasingly common assumption of theories of memory is that the meaning of a word can be represented by a vector which places a word as a point in a multidimensional semantic space (Bower, 1967; Landauer & Dumais, 1997; Lund & Burgess, 1996; Morton, 1970; Norman, & Rumelhart, 1970; Osgood, Suci, & Tannenbaum, 1957; Underwood, 1969; Wickens, 1972). The main requirement of such spaces is that words that are similar in meaning are represented by similar vectors. Representing words as vectors in a multidimensional space allows simple geometric operations such as the Euclidian distance or the angle between the vectors to compute the semantic (dis)similarity between arbitrary pairs or groups of words. This representation makes it possible to make predictions about performance in psychological tasks where the semantic distance between pairs or groups of words is assumed to play a role.

One recent framework for placing words in a multidimensional space is Latent Semantic Analysis or LSA (Derweester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). The main assumption is that similar words occur in similar contexts with context defined as any connected set of text from a corpus such as an encyclopedia, or samples of texts from textbooks. For example, a textbook with a paragraph about "cats" might also mention "dogs", "fur", "pets" etc. This knowledge can be used to assume that "cats" and "dogs" are related in meaning. However, some words are clearly related in meaning such as "cats" and "felines" but they might never occur simultaneously in the same context. Such words are related primarily through indirect links because they share similar contexts. The technique of singular value decomposition (SVD) can be applied to the matrix of word-context co-occurrence statistics. In this procedure, the direct and indirect relationships between words and contexts in the matrix are analyzed with simple matrix-algebraic operations and the result is a high dimensional space in which words that appear in similar contexts are placed in similar regions of the space. Landauer and Dumais (1997) applied the LSA approach to over 60,000 words appearing in over 30,000 contexts of a large encyclopedia. More recently, LSA was applied to over 90,000 words appearing in over 37,000 contexts of reading material that an English reader might be exposed to from 3rd grade up to 1st year of college from various sources such as textbooks, novels, and newspaper articles. The SVD method placed these words in a high dimensional space with the number of dimensions

chosen between 200 and 400. The LSA representation has been successfully applied to multiple choice vocabulary tests, domain knowledge tests and content evaluation (see Landauer & Dumais, 1997; Landauer et al. 1998).

Another framework that has been used to place words in a high dimensional semantic space is the Hyperspace Analogue to Language (Burgess, Livesay, & Lund, 1998; Lund & Burgess, 1996; see Burgess & Lund, 2000 for an overview). The HAL model develops high dimensional vector representations for words based on a co-occurrence analysis of large samples of written text. For 70,000 words, the co-occurrence statistics were calculated in a 10 word window that was slid over the text from a corpus of over 320 million words (gathered from Usenet newsgroups). For each word, the co-occurrence statistics were calculated for the 70,000 words appearing before and after that word in the 10 word window. The resulting 140,000 values for each word were the feature values for the words in the HAL representation. Because the representation is based on the context in which words appear, the HAL vector representation is also referred to as a contextual space: words that appear in similar contexts are represented by similar vectors. The HAL and LSA approach are similar in one major assumption: similar words occur in similar contexts. In both HAL and LSA, the placement of words in a high dimensional semantic space is based on an analysis of the co-occurrence statistics of words in their contexts. In LSA, a context is defined by a relatively large segment of text whereas in HAL, the context is defined by a window of 10 words.

LSA and HAL are both corpus based methods that contrast sharply with older methods of constructing semantic spaces that depend on human judgments such as pairwise similarity ratings[1]. In this method, participants rate the semantic similarity for pairs of words. Then, those similarity ratings are subjected to multidimensional scaling analyses to derive vector representations in which similar vectors represent words similar in meaning (Caramazza, Hersch, & Torgerson, 1976; Rips, Shoben, & Smith, 1973; Schwartz & Humphreys, 1973). While this method is straightforward and has led to interesting applications (e.g. Caramazza et al; Romney, Brewer, & Batchelder, 1993), it is clearly impractical for large number of words as the number of ratings that must be collected goes up quadratically with the number of stimuli. One great advantage of LSA and HAL over approaches depending on pairwise similarity ratings is that almost any number of words can be placed in a semantic/contextual space. This is possible because these methods rely uniquely on samples of written text (of which there is a virtually unlimited amount) as opposed to ratings provided by participants.

In this research, we will introduce a related but new method for creating psychological spaces that is based on an analysis of a large free association database collected by Nelson, McEvoy, and Schreiber (1999) containing norms for first associates for over 5000 words. This method places over 5000 words in a psychological space that we will call Word Association Space (WAS). We believe such a construct will be very useful in the modeling of episodic memory phenomena. At present it is not clear how well episodic memory performance in recognition or recall is predicted by LSA or HAL. We do know that word associations play an important role in episodic memory since it has been shown that the associative structure of words plays a central role in recall (e.g. Bousfield, 1953; Cramer, 1968; Deese, 1959a,b, 1965; Jenkins, Mink, & Russell, 1958), cued recall (e.g. Nelson, Schreiber, & McEvoy, 1992) and priming (e.g. Canas, 1990; see also Neely, 1991), and recognition (e.g. Nelson, Zhang, & McKinney, submitted). For example, Deese (1959a,b) found that the inter-item associative strength of the words in a study list can predict the number of words recalled, the number of intrusions, and the frequency with which certain words intrude. Based on the classic experiments by James Deese (1959b), Roediger and McDermott (1995) revived interest in the paradigm that is now known as the false memory paradigm (e.g. Brainerd, & Reyna, 1998; Brainerd, Reyna, & Mojardin, 1999; Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996; Tussing & Green, 1997; Shiffrin, Huber, & Marinelli, 1995). In the typical false memory experiment, participants study words that are all semantically related to a non-studied critical word. In a subsequent recognition test, the critical word typically lead to a higher false alarm rate than that for unrelated foils (and sometimes quite high in comparison to that for studied words). In a free recall test, participants falsely intrude the critical word at a rate higher than unrelated words (and sometimes at rates approaching those for studied words). These studies show that episodic memory can be strongly influenced by semantic similarity.

In the present research, we will compare the performance of LSA with WAS in three episodic memory tasks: recognition memory, free recall and cued recall[2]. It was expected that the similarity structure in WAS would be well suited to predict various semantic similarity effects in these episodic memory tasks. To further our understanding of the similarity structure of WAS, we performed several analyses. First, we compared the predictions of WAS and LSA for the strength of the associates obtained in

the free association task. Because WAS is explicitly based on the free association norms, it was expected that there would be a close relationship between distance in the psychological space and associative strength in free association. In another analysis, the issue of whether WAS captures semantic or associative relationships (or both) in predicting similarity ratings is addressed. It will be argued that it is difficult to make a categorical distinction between purely associative and purely semantic relationships. Finally, in the last analysis we analyze the ability of WAS to predict asymmetries in the direction of association. Because free association norms contain many examples of words that are strongly associated in one but not the other direction, predicting these asymmetries with a model such as WAS presents an interesting challenge given that the similarity between two words is defined to be symmetric.

Word Association Spaces

Deese (1962,1965) asserted that free associations are not the result of haphazard processes and that they arise from an underlying regularity in pre-existing associative connections. He laid the framework for studying the meaning of linguistic forms that can be derived by analyzing the correspondences between distributions of responses to free association stimuli: "The most important property of associations is their structure - their patterns of intercorrelations" (Deese, 1965, p.1). Deese (1962, 1965) applied factor analyses to the overlap in the distribution of free association responses for a small set of words and argued that these analyses could be used to learn about the mental representation of words. In this paper, we capitalized on Deese's ideas of utilizing the pattern of intercorrelations in the free association norms by placing a large number of word associations in a semantic space and then used them to predict semantic similarity effects in memory. Instead of factor analyses, we used the techniques of singular value decomposition (SVD) and metric multidimensional scaling analyses that are closely related to factor analysis (e.g., see Johnson & Wichern, 1998) but that can be applied to large volumes of data.

The SVD method has been successfully applied in LSA (e.g. Landauer & Dumais, 1997) to uncover the patterns of intercorrelations of co-occurrence statistics for words appearing in contexts. In this research, the SVD method was applied on a large database of free association norms collected by Nelson et al. (1999) containing norms of first associates for over 5000 words. The result is that the words are placed in a high dimensional semantic space. Because of the SVD method, the similarity between words in this psychological space is best expressed by the cosine of the angle between the two corresponding vectors (see Derweester et al. 1990). We also performed additional analyses by applying metric multidimensional scaling (MDS) methods (see Torgeson, 1952; Schiffman, Reynolds & Young, 1981) on the norms such that the Euclidian distance

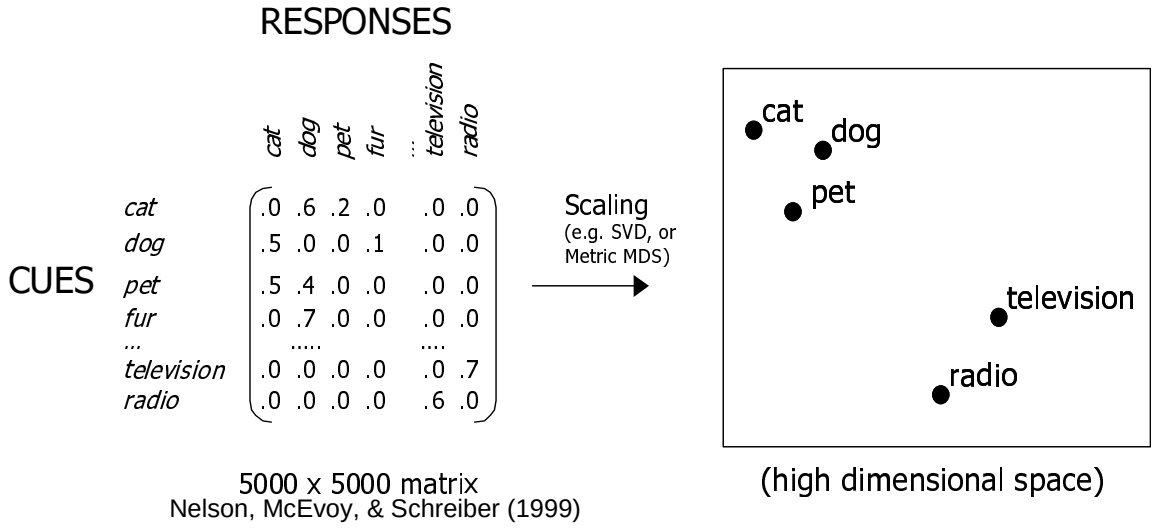

Figure 1. Illustration of the creation of Word Association Spaces (WAS). By scaling the word associations of a large database of free association norms, words are placed in a high dimensional semantic space. Words with similar associative relationships are placed in similar regions of the space.

Table 1. Overview of Methods for Quantifying Associative/Semantic similarity

Method	Based on:	(Dis)similarity Measure	Vocabulary Size
Associations	$S^{(1)}$: forward plus backward strength	$S^{(1)}$	5000+
	$S^{(2)}$: forward plus backward plus two step associative strengths	$S^{(2)}$	5000+
WAS	svd of $S^{(1)}$	cosine of angle	2500
	svd of $S^{(2)}$	cosine of angle	2500, 5000+
	metric mds of T (see text)	Euclidian distance	
LSA	svd of word-context matrix of encyclopedia	cosine of angle	60,000+
	svd of word-context matrix of tasa documents	cosine of angle	90,000+

between two words in the resulting psychological space captures the dissimilarity between two words. We will refer to the general method of placing the words in a space as well as the space itself as Word Association Space (WAS). For an overview of all the different methods to quantify associative/semantic similarity in this research, see Table 1.

By applying scaling methods such as SVD and metric MDS on the norms, we hope to uncover the latent information available in the free association norms that is not directly available by investigating simple measures for associative strengths based on the direct and indirect associative strengths through short chains of associates (e.g., Nelson & Zhang, 2000). The basic approach is illustrated in Figure 1. The free association norms were represented in matrix form with the rows representing the cues and the columns representing the responses. The entries in the matrix are filled by some measure of associative strength between cues and responses. By applying scaling methods on the matrix, words are placed in a high dimensional space such that words with similar associative patterns are placed in similar regions of the space.

In total, more than 6000 people participated in the collection of the free association database of Nelson et al. (1999). An average of 149 (SD = 15) participants were each presented with 100-120 English words. These words served as cues (e.g. "cat") for which participants had to write down the first word that came to mind (e.g. "dog"). For each cue the proportion of subjects that elicited the response to the cue was calculated (e.g. 60% responded with "dog", 15% with "pet", 10% with "tiger", etc).

Scaling by Singular Value Decomposition

We will first explain how SVD is applied to the norms and then continue to the metric MDS method. The method of SVD can be applied to any matrix containing some measure of strength or co-occurrence between two words. Although many different ways have been proposed to calculate an index of associative strength between two words (e.g., Marshall & Cofer, 1963; Nelson & Zhang, 2000), we will restrict ourselves to two simple measures of associative strength. Let A_{ij} represent the proportion of subjects that gave the response j to the cue i. The simplest measure would be to take A_{ij} itself. In the norms, the associative strengths A_{ij} are often highly asymmetric where the associative strength in one direction is strong while it is weak or zero in the other direction. Even though SVD can be easily applied to asymmetric matrices, the results are more interpretable when it is applied to symmetric matrices[3]. In a later section in the paper, we will show that symmetrizing the norms does not necessarily mean that asymmetries in the word associations cannot be predicted. Therefore, in our first measure for associative strength we take:

$$S_{ij}^{(1)} = A_{ij} + A_{ji}$$

$S^{(1)}_{ij}$ is equivalent to adding forward strength to backward strength. This measure is of course symmetric so that $S^{(1)}_{ij} = S^{(1)}_{ji}$. This measure is indexed by (1) because it based on only the direct association between i and j and involves only one associative step going from i to j. In the norms of

Nelson et al. (1998), subjects were only allowed to give the first response that came to mind. The second strongest response in one subjects' mind might be elicited by another subject or it might not be elicited at all if the first response is a strong associate. Therefore, the $S^{(1)}$ measure might be underestimating the associative strength between two words especially in cases where the measure is zero (Nelson et al., 1998). In the second measure for associative strength, we take:

$$S_{ij}^{(2)} = S_{ij}^{(1)} + \sum_k S_{ik}^{(1)} S_{kj}^{(1)}$$

This equals the forward plus backward plus mediated strength through other associates. Note that this measure involves the direct strength between i and j as well as the indirect strength by summing over all paths from i to k to j, the product of the symmetric associative strengths between i and k, and k and j. These indirect associative strengths involve the two step probabilities of going from i to j and vice versa and hence the index (2). Research has shown that the indirect associative strengths play a role in cued recall (Nelson, Bennet, & Leibert, 1997; Nelson & Zhang, 2000) and recognition (Nelson, Zhang, & McKinney, submitted). For example, Nelson & Zhang (2000) found that including the indirect associative strengths in a measure for associative strength significantly increases the explained variance in the extra-list cued recall task.

We applied SVD separately on these two measures of associative strength. The result of each SVD is the placement of words in a high dimensional space, so that words that have similar associative structures are represented by similar vectors. Because of the SVD method, and based on work in LSA (see Derweester et al., 1990), a suitable measure for the similarity between two words is the cosine of the angle between two word vectors. Let $\vec{X}_i$ represent the vector in WAS for word i. The similarity between words i and j is calculated by:

$$similarity(i, j) = \cos(\alpha) = \frac{\vec{X}_i \cdot \vec{X}_j}{\left\|\vec{X}_i\right\| \left\|\vec{X}_j\right\|}$$

where $\left\|\vec{X}\right\|$ is the length of the vector and $\vec{X}_i \cdot \vec{X}_j$ represents the inner product between vectors i and j. Two words that are similar in meaning or that have similar associative structures are expected to have high similarity as defined by the cosine of the angle between the two word vectors. The SVD of the associative strengths can uncover the latent relationships between words. In the SVD of $S^{(1)}$, words that are not direct associates of each other can still be represented by similar vectors if their associates are related. In the SVD of $S^{(2)}$, words that not directly associated or indirectly associated through one intermediate associate, can still be represented by similar vectors if the associates of the associates of the words are related. In other words, the whole pattern of direct and indirect correlations between associations is taken into account when placing words in the semantic space.

An important variable (which we will call k) is the number of dimensions of the space. One can think of k as the number of feature values for the words. We varied k between 10 and 500. The number of dimensions will determine how much the information of the free association database is compressed[4]. With too few dimensions, the similarity structure of the resulting vectors does not capture enough detail of the original associative structure in the database. With too many dimensions or the number of dimensions approaching the number of cues, the information in the norms is not compressed enough so that we might expect that the similarity structure of the vectors does not capture enough of the indirect relationships in the associations between words. In the analyses of predicting performance in a variety of tasks (recognition, free and cued recall), we will show that although the optimal value of k depends on the specific task, intermediate values of k between 200 and 500 are appropriate for this method.

Scaling by Metric-MDS

An interesting comparison for the two WAS spaces based on SVD would be to construct a metric space in which the distance between two words, i.e., their dissimilarity, can be measured by the Euclidian distance between their vectors. Metric MDS is a classic method for placing stimuli in a space such that the Euclidian distance between points in the space approximates the Euclidian distances in the dissimilarity matrix (see Torgeson, 1952; Schiffman, Reynolds & Young, 1981). In order to apply metric MDS, estimates are needed for the distance between any two words. In fact, all non-diagonal entries in the matrix have to be filled with some estimate for the distance between words since no missing values are allowed in the method[5]. This raises the problem how to estimate the distance between i and j when the associative strength as measured by $S_{ij}^{(1)}$ is zero[6].

In our solution of this problem, we were inspired by network models for proximity data (e.g. Cooke, Durso, & Schvaneveldt, 1986; Klauer, & Carroll, 1995). In these network models, dissimilarity between two stimuli is calculated by the shortest path

between two nodes in a graph. In this research, we can use the word association norms as defining a graph: two words are linked by an edge if they have nonzero associative strengths. We will use the symmetric $S^{(1)}$ associative strengths because in the graph defined by $S^{(1)}$, it is possible to reach any word from any other word in the graph (in fact, the maximum number of steps between any pair of words is four). The distance between two words will be defined as the negative logarithm of the product of the associative strengths along the shortest path in the network defined by $S^{(1)}$. This is equivalent to the (negative) sum of the logs of the associative strengths along the shortest path:

$$T_{ij} = -\log\left(S^{(1)}_{ik} S^{(1)}_{kl} \cdots S^{(1)}_{qj}\right) = -\left[\log S^{(1)}_{ik} + \log S^{(1)}_{kl} + ... + \log S^{(1)}_{qj}\right]$$

Here, the shortest path between words i and j is from i to k to l through other words to q and finally j. With this distance measure, word pairs with weak or long associative paths are assigned large distances whereas word pairs with short or strong associative paths are assigned small distances. The distances T_{ij} were calculated for all word pairs in the word association database. Then, these distances were scaled by metric-MDS. The result is that the words are placed in a multidimensional space and the dissimilarity or distance between two words is expressed by the Euclidian distance between the two corresponding word vectors:

$$distance(i, j) = \left[\sum_k \left(X_{ik} - X_{jk}\right)^2\right]^{1/2}$$

Because of computational constraints, it was not possible to apply metric-MDS to the full matrix T containing the distances for all word pairs. Instead, we chose 2500 words from the original 5018 words in the word association database. The words in this smaller set included words appearing in various experiments listed in the next section and included a selection of randomly chosen words from the original set. The SVD procedure was applied on both the smaller matrix of 2500 words as well as the set of 5018 words.

As with the SVD procedure, the choice of the number of dimensions in the space is important. Having too few dimensions or too many dimensions might lead to suboptimal performance when predicting performance in various memory tasks. As with the SVD scaling procedure, the number of dimensions was varied between 10 and 500.

Predicting Semantic Similarity Effects in Memory

Since Deese's (1959b) classic study on intrusions in free recall, many studies have shown that memory errors are in part based on semantic overlap between the response and the contents of memory. In this research, we introduced WAS as a way of quantifying the semantic similarity between words that might help in predicting these memory errors. The data from a recognition experiment, Deese's original free recall experiment and a cued recall experiment were taken as a basis for testing various models that capture semantic similarity. We tested three WAS based measures for semantic similarity. The first two were based on the SVD of $S^{(1)}$, the one step symmetric associative strengths, and on the SVD of $S^{(2)}$, the one plus the two step associative strengths involving indirect associative strengths. In these two semantic spaces (as well as in LSA) the cosine of the angle between two words expresses the similarity between two words. The last WAS measure was based on metric-MDS of the shortest path associative strengths. In this space, the Euclidian distance between two word vectors is taken as a measure for the dissimilarity between two words. These WAS scaling solutions were contrasted with the (unscaled) associative strengths $S^{(1)}$ and $S^{(2)}$ that were taken as control comparisons. We also tested two LSA based measures, one was based on a corpus of an encyclopedia and another on a corpus called tasa that included reading material that an English reader might be exposed to from 3rd grade up to 1st year of college. The different methods that are compared are listed in Table 1.

Recognition Memory: Semantic Similarity Ratings

In a study by the first two authors (Steyvers & Shiffrin, submitted: Experiment 1), 89 subjects studied 144 words that contained 18 semantic categories of 5 words each. Based on a study by Brainerd and Reyna (1998), subjects gave two ratings for each of 100 test items. In one rating, they were instructed to judge whether the item was old or new and were told to judge semantically similar distractors as "new". In another rating, they were instructed to rate (on a six point scale) how semantically similar the item was to the studied items. We focused on the semantic similarity ratings for the new items from this study. For each subject, the 72 new test items were randomly selected from a larger pool of 144 words. An average of 44 (SD=4.87) subjects rated the semantic similarity for each of the 144 words that might appear as new

words in the test list[7]. The semantic similarity ratings are theoretically interesting because they can be used to test models of semantic similarity. Subjects merely have to remember how similar the item was to the studied items without being forced to give old-new judgments that might be more influenced by various strategic retrieval factors (such as word frequency or previous retrievals).

Many memory models assume that a recognition memory judgment is produced by calculating the global familiarity involving the summed similarity between the test item and the episodic traces in memory (e.g. Minerva II, Hintzman 1984, 1988; SAM, Gillund & Shiffrin, 1984). More recently, Shiffrin and Steyvers (1997, 1998) and McClelland & Chappell (1998) have proposed recognition memory models that produce recognition judgments with Bayesian decision processes. McClelland & Chappell (1998) proposed that the best match (i.e., maximum similarity) between the test item and the episodic traces in memory forms the basis for the recognition judgment. Shiffrin & Steyvers (1998) showed that in the Bayesian framework, a maximum similarity process produced results very similar to a

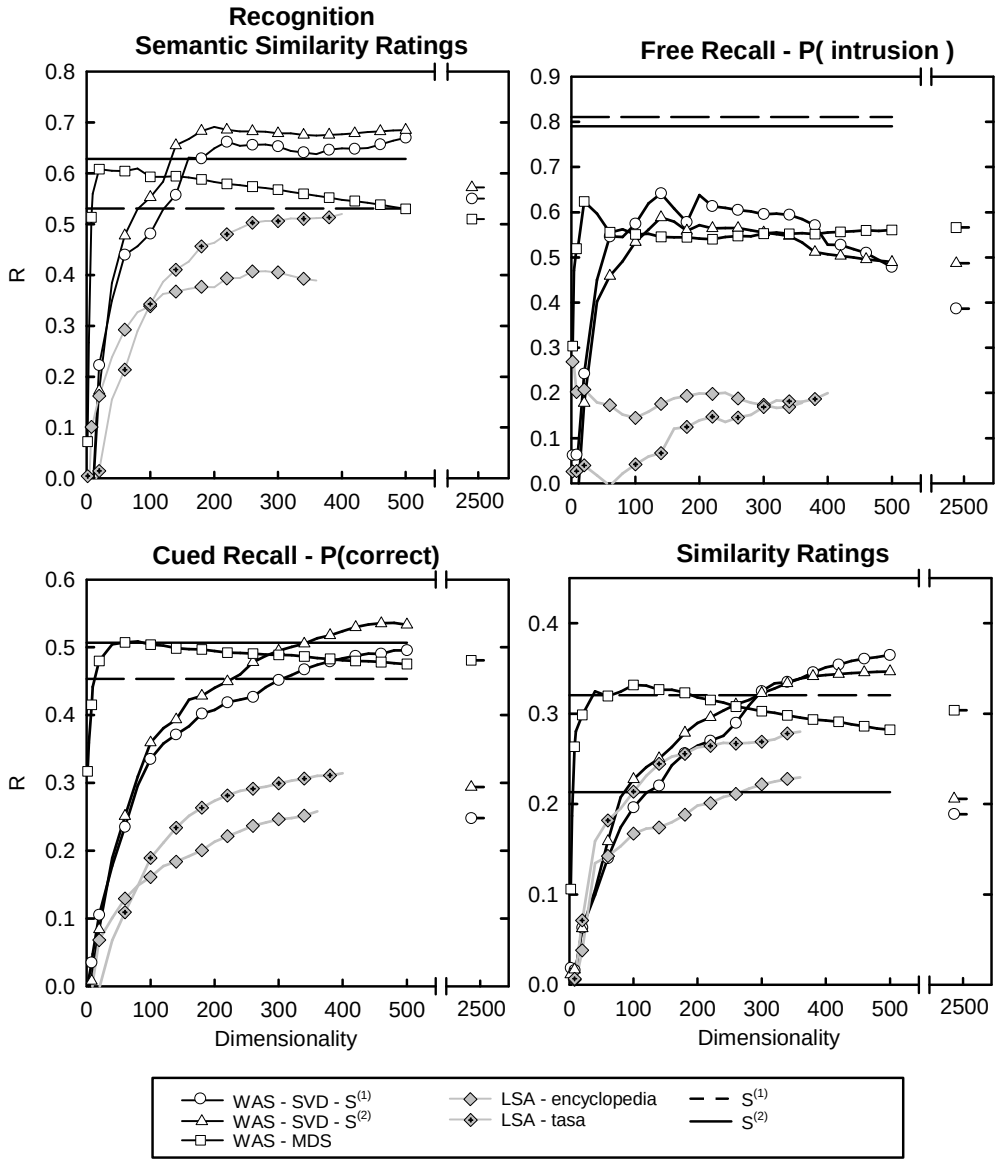

Figure 2. Correlations of different measures of semantic similarity for different dimensionalities. Data are taken from recognition memory, cued recall, free recall, and similarity ratings experiments. See text for details.

summed similarity process. In this research, our aim is not to test these models specifically but to use and simplify the underlying mechanisms to predict semantic similarity ratings.

Based on the global familiarity and Bayesian recognition memory models, the semantic similarity ratings in the recognition memory experiment should have been correlated with the sum or maximum of the similarity between the test item and all study words. To facilitate comparisons, all expected negative correlations were converted to positive correlations. Because the results were very similar for the sum and maximum calculations, we will list only the results for the maximum calculation.

The top left panel of Figure 2 shows the correlations between maximum similarity and number of dimensions (10-500) for the three WAS and two LSA based measures. For the SVD based semantic spaces, increasing the number of dimensions in either WAS or LSA increases the correlation generally up to around 200-300 dimensions. For WAS, an additional data point was plotted for 2500 dimensions which is the maximum number of dimensions given that the matrix contained only 2500 words. Although a space was available for all 5018 words in the free association database, the vocabulary was restricted to 2500 words in order to able to compare it to the metric MDS solutions (see previous section). This data point for 2500 dimensions was included because it represents the case where none of the indirect relationships in word association matrix are exploited and as such, no dimensionality reduction is performed. As can be observed, the correlation is lower for 2500 dimension indicating that some dimensionality reduction is needed to predict the semantic similarity ratings. Also, the SVD based on $S^{(2)}$ led to better correlations than the SVD based on $S^{(1)}$. This implies that adding the indirect associations in a measure for associative strength helps in predicting recognition memory performance. The two horizontal lines in the plot indicate the correlation when the associative strengths $S^{(1)}$ and $S^{(2)}$ are used as a measure for semantic similarity. The correlation is higher for $S^{(2)}$ than $S^{(1)}$ which again implies that in recognition memory, the indirect associative strengths help in predicting performance. Interestingly, the SVD scaling of $S^{(2)}$ gave higher correlations than associative strengths $S^{(2)}$ themselves. Even though $S^{(2)}$ includes the forward, backward and all two step associative strengths, applying the SVD and reducing the redundancies in the matrix of $S^{(2)}$ helped to increase the correlation. In other words, the indirect relationships and patterns of correlations that go beyond those of the two step associative strengths were utilized by the SVD procedure and these were beneficial in predicting the ratings from this recognition memory experiment.

The metric-MDS solution shows quite a different pattern of results than the SVD solution. The best correlation was obtained with 20-40 dimensions which is much lower than the number of dimensions typically needed in the SVD solutions of either WAS or LSA. Although the best correlation for metric-MDS was 0.6 as opposed to 0.7 for the SVD based solutions, it is interesting that relatively good performance can be achieved in semantic spaces that are of low dimensionality. Although specifying why this effect occurs is outside the scope of this paper, it could be related to the estimates involving the shortest associative path between words. As described in the previous section, in order to apply metric-MDS, estimates were needed for the distances between all word pairs in the vocabulary. The shortest associative path distance was proposed to meet this requirement; estimates were even generated for word pairs that were not associated directly or even indirectly through a chain of two associates. In SVD, no such estimates are required and those entries were left at zero. It is possible then, that the filling in process of all word pair dissimilarities by the shortest associative path distances helped in the global placement of all words in the semantic space.

Of the two corpera in LSA, the tasa corpus led to much better performance than the encyclopedia corpus. This difference is not surprising since the tasa corpus includes material that reflects much more closely the reading material an English reader is exposed to which in turn might lead to semantic spaces that are more psychologically plausible in terms of predicting semantic similarity effects in recognition memory. Comparing WAS to LSA, it becomes clear that WAS leads to much higher correlations than LSA. We will leave the interpretation of this finding for the discussion.

Predicting Extralist Cued Recall

In extra-list cued recall experiments, after studying a list of words, subjects are presented with cues that can be used to retrieve words from the study list. The cues themselves are novel words that were not presented during study, and typically each word is associatively and/or semantically related to one of the studied words. The degree to which a cue is successful in retrieving a particular target word is a measure of interest because this might be related to the associative/semantic overlap between cues and their targets. Research in this paradigm (e.g., Nelson & Schreiber, 1992; Nelson, Schreiber, & McEvoy, 1992; Nelson, McKinney, Gee, & Janczura, 1998; Nelson & Zhang, 2000) has shown that the associative strength between cue and target is one

important predictor for the percentage of correctly recalled targets. Therefore, we expect that the WAS similarity between cues and targets are correlated with the percentages of correct recall in these experiments. We used a database containing the percentages of correct recall for 1115 cue-target pairs[8] from over 29 extralist cued recall experiments from Doug Nelson's laboratory (Nelson, 2000; Nelson & Zhang, 2000).

The correlations between the various measures for semantic similarity and the observed percentage correct recall rates are shown in the bottom left panel of Figure 2. Overall, the results are very similar to the results obtained for the recognition memory experiment. The WAS space based on $S^{(2)}$ led to better performance than the WAS space based on $S^{(1)}$. Also, the associative strengths $S^{(2)}$ leads to better performance then the $S^{(1)}$ associative strengths. These findings are consistent with findings by Nelson & Zhang (2000) that show that the indirect relationships in word association norms can help in predicting cued recall performance. Interestingly, the plot also shows that the WAS space based on $S^{(2)}$ does somewhat better than the associative strengths $S^{(2)}$ it was based on. This advantage implies that applying dimensionality reduction to make greater use of the indirect associative connections helped in predicting cued recall. Finally, as with the recognition results, the WAS space correlates better with cued recall than LSA.

Predicting Intrusion Rates in Free Recall

In a classic study by Deese (1959b), the goal was to predict the intrusion rates of words in free recall. Fifty participants studied the 12 strongest associates to each of 36 critical lures while the critical lures themselves were not studied. In a free recall test, some critical lures (e.g. "sleep") were falsely recalled about 40% of the time while other critical lures (e.g. "butterfly") were never falsely recalled. Deese was able to predict the intrusion rates for the critical lures on the basis of the average associative strength from the studied associates to the critical lures and obtained a correlation of R=0.80. Because Deese could predict intrusion rates with word association norms, the WAS vector space derived from the association norms should also predict them. Critical items with high average similarity (or low average distance) to the list words in the semantic space should be more likely to appear as intrusions in free recall. The average similarity (average distance) was computed between each critical lure vector and list word vectors, and the correlations were computed between these similarities and observed intrusion rates.

The top right panel in Figure 2 shows the results. The pattern of results is quite different than the pattern of results for either recognition or cued recall. The best correlation of 0.82 was obtained with $S^{(1)}$, the sum of backward and forward associative strength. This result is very similar to the correlation of 0.80 Deese obtained with his word association norms. Interestingly, the plot shows that any manipulation that includes the indirect associations leads to worse performance than using the direct associations only. The WAS space based on $S^{(2)}$ now does worse than the WAS space based on $S^{(1)}$, and either space correlates more poorly than when using the associative strengths $S^{(1)}$ and $S^{(2)}$ themselves.

These findings imply that direct associative strengths are the best predictors of intrusion rates in free recall. One explanation for this finding is related to implicit associative responses (IAR's). Underwood (1969) has argued that during study, the words associated with the study words are thought of and might be stored in memory as an implicit associative response. In Deese's study, it is likely that IAR's were generated because the critical lures were all strongly associated to the list words. Therefore, during recall, the words that were actually presented and words that were thought of during study might be confused leading in some cases to dramatic intrusion rates. Because free associations measure what responses are thought of given specific cues, the direct associative strengths can be argued to be good predictors of the strength of implicit associative responses and subsequent intrusion rates.

Similarities and differences between WAS and free association norms

Because WAS places words in a multidimensional space based on the pattern of inter-correlations in free association, the similarities and differences of WAS and the free association norms needs to be determined. In the previous section, it was established that in recognition and cued recall, WAS leads to somewhat better correlations with observed results than the associative strengths it was based on. In this section, the similarity structure of WAS is investigated and compared to the free association norms. First, we investigate the degree to which the neighborhood similarity structure in WAS can be used to predict the order of response strengths in free association. Then, we address the issue of what kind of relationship WAS captures: semantic, associative or both. Finally, we assess whether asymmetries in the direction of association can be predicted by the neighborhood similarity structure in WAS.

Predicting the Output Order of Free Association Norms

Because the word vectors in WAS are based explicitly on the free association norms, it is of interest to check whether the output order of responses (in terms of associative strength) can be predicted by WAS. To simplify the analysis, the results are only presented for WAS based on $S^{(2)}$. We took this space because it performed well in all three memory tasks and because we had a solution available for all 5000+ words appearing in the free association norms (the metric space was limited to a vocabulary of 2500 words).

Table 2. Median Rank of the Output-order in WAS and LSA of Response Words to Given Cues for the 10 Strongest Responses in the Free Association Norms.

	Rank of Response in Free Association									
k	1	2	3	4	5	6	7	8	9	10
Word Association Space (WAS)										
10	86	187	213	249	279	291	318	348	334	337
50	13	36	49	62	82	98	106	113	125	132
100	6	17	26	36	43	62	65	73	78	85
200	3	8	15	20	28	39	40	48	56	58
300	2	6	12	16	21	31	35	38	43	49
400	1	5	10	14	19	27	32	35	38	44
Latent Semantic Analysis (LSA)										
10	733	798	846	845	922	897	903	920	939	955
50	231	313	371	422	475	494	526	510	559	583
100	115	193	256	307	359	384	411	413	451	463
200	63	125	185	225	285	319	347	339	389	395
300	46	99	159	197	254	294	321	324	375	374
400	37	90	149	185	239	278	310	308	349	366

Note: In these analyses, WAS was based on the svd of $S^{(2)}$ and LSA was based on the tasa corpus

We took the 10 strongest responses to each of the cues in the free association norms and ranked them according to associative strengths. For example, the response 'crib' is the 8th strongest associate to 'baby' in the free association norms, so 'crib' has a rank of 8 for the cue 'baby'. Using the vectors from WAS, the rank of the similarity of a specific cue-response pair was computed by ranking the similarity among the similarities of the specific cue to all other possible responses. For example, the word 'crib' is the 2nd closest neighbor to 'baby' in WAS, so 'crib' has a rank of 2 for the cue 'baby'. In this example, WAS has put 'baby' and 'crib' closer together than might be expected on the basis of free association norms. Averaged across the words in the corpus, Table 2 gives for each of the first ten ranked responses in free association (the columns) the median rank in WAS. The median was used to avoid excessive skewing of the average by a few high ranks. An additional variable that was tabulated in Table 2 is k, the number of dimensions of WAS.

There are three trends to be discerned in Table 2. First, it can be observed that for 400 dimensions, the strongest responses to the cues in free association norms are predominantly the closest neighbors to the cues in WAS. Second, responses that have higher ranks in free association have on average higher ranks in WAS. However, the output ranks in WAS are in many cases far higher than the output ranks in free association. For example, with 400 dimensions, the third ranked response in free association has a median rank of 10 in WAS. Third, for smaller dimensionalities, the difference between the output order in free association and WAS becomes larger.

To summarize, given a sufficiently large number of dimensions, the strongest response in free association is represented (in most cases) as the closest neighbor in WAS. The other close neighbors in WAS are not necessarily associates in free association (at least not direct associates). We also analyzed the correspondence between the similarities in the LSA space (Landauer & Dumais, 1997) based on the tasa corpus with the order of output in free association. As can be observed in Table 2, the rank of the response strength of the free association norms clearly has an effect on the ordering of similarities in LSA: strong associates are closer neighbors in LSA than weak associates. However, the overall correspondence between predicted output ranks in LSA and ranks in the norms is weak. The overall weaker correspondence between the norms and similarities for the LSA approach than the WAS approach highlights one obvious difference between the two approaches. Because WAS is based explicitly on free association norms, it is expected and shown here that words that are strong associates are placed close together in WAS whereas in LSA, words are placed in the semantic space in a way more independent from the norms.

To get a better idea of the kinds of neighbors words have in WAS, in Table 3, we list the first five

Table 3. The Five Nearest Neighbors in WAS for the First 40 cues in the Russell & Jenkins (1954) Norms.

	Neighbor				
Cue	1	2	3	4	5
afraid	scare(1)[7]	fright(4)[14]	fear(2)[1]	scared[2]	ghost(5)[106]
anger	mad(1)[1]	Angry	rage(5)[4]	enrage	fury[21]
baby	child(1)[2]	crib(8)[13]	infant(6)[7]	cradle	diaper(13)
bath	clean(2)[1]	soap(7)[3]	water(3)[2]	dirty[7]	suds[49]
beautiful	pretty(1)[2]	ugly(2)[1]	cute[39]	girl(4)	flowers[10]
bed	sleep(1)[1]	tired(11)[13]	nap	rest[5]	doze
bible	god(1)[1]	church(3)[3]	religion(4)[4]	Jesus(5)[8]	book(2)[2]
bitter	sweet(1)[1]	sour(2)[2]	Candy	lemon(5)[7]	chocolate[4]
Black	white(1)[1]	Bleach	color(3)[7]	dark(2)[2]	minority
blossom	flower(1)[1]	petals[46]	Rose(5)[7]	tulip	daisy
blue	color(5)[4]	red(3)[2]	Jeans	crayon	pants
boy	girl(1)[1]	Guy	Man(4)[2]	woman	nephew[54]
bread	butter(1)[1]	toast[19]	rye[26]	loaf(3)[16]	margarine
butter	bread(1)[1]	toast(6)[18]	rye	peanut	margarine(2)[34]
butterfly	bug(15)[10]	insect(6)[2]	fly(4)[5]	roach[76]	beetle
cabbage	green(4)[7]	food(10)[4]	vegetable(2)[3]	salad(12)[5]	vegetables
carpet	floor(2)[2]	tile(15)	rug(1)[1]	ceiling	sweep[14]
chair	table(1)[1]	seat(4)[4]	sit(2)[2]	couch(3)[20]	recliner
cheese	cracker(2)	cheddar(6)[23]	Swiss(7)[19]	macaroni[39]	pizza
child	baby(1)[1]	kid(2)[7]	adult(3)[3]	young(8)[6]	parent(6)[11]
citizen	person(1)[3]	country(3)[5]	people[7]	flag[12]	American(2)[4]
city	town(1)[1]	state(2)[3]	country(9)[4]	New York(4)	Florida
cold	hot(1)[1]	ice(2)[5]	warm(6)[3]	chill	pepsi
comfort	chair(3)[1]	Table	seat	couch(2)[26]	sleep[7]
command	tell(4)[7]	army(5)[2]	rules	navy[17]	ask[22]
cottage	house(1)[1]	home(4)[4]	cheese(2)[3]	cheddar	Swiss
dark	light(1)[1]	Bulb	night(2)[2]	lamp	day
deep	water(3)[3]	ocean(2)[6]	faucet	pool[53]	splash
doctor	nurse(1)[1]	physician(5)[15]	surgeon(6)	medical[83]	stethoscope[21]
dream	sleep(1)[1]	fantasy(4)[19]	bed[7]	nap	tired[92]
eagle	bird(1)[1]	Chirp	blue jay	nest(10)[5]	sparrow[30]
earth	planet(2)[8]	mars[14]	Jupiter[97]	Venus[50]	Uranus
eating	food(1)[1]	eat[30]	hungry(3)[4]	restaurant[75]	meal[30]
foot	shoe(1)[1]	sock[16]	toe(2)[3]	sneaker	leg(5)[4]
fruit	orange(2)[3]	apple(1)[1]	juice(9)[12]	citrus[35]	tangerine[55]
girl	boy(1)[1]	guy(6)	man[9]	woman(3)[2]	pretty(4)[6]
green	grass(1)[1]	lawn[41]	cucumber	vegetable[76]	spinach[76]
hammer	nail(1)[1]	tool(2)[7]	wrench	screwdriver	pliers[21]
hand	finger(1)[2]	arm(3)[3]	foot(2)[1]	leg(13)[11]	glove(4)[4]
hard	soft(1)[1]	easy(3)[3]	difficult[19]	difficulty	simple

Note: numbers in parentheses and square brackets indicate ranks of responses in norms of Nelson et al. (1998) and Russell & Jenkins (1954) respectively.

neighbors in WAS (using 400 dimensions) to 40

cue words taken from Russell and Jenkins (1954).

For all neighbors listed in the table, if they were associates in the free association norms of Nelson et al., then the corresponding rank in the norms is given in parentheses and those from Russell and Jenkins are shown in brackets. The comparison between these two databases is interesting because Russell and Jenkins allowed participants to generate as many responses they wanted for each cue while the norms of Nelson et al. contain first responses only. We suspected that some close neighbors in WAS are not direct associates in the Nelson et al. norms but that they would have been valid associates if participants had been allowed to give more than one association per cue. In Table 4, we list the percentages of neighbors in WAS of the 100 cues of the Russell and Jenkins norms (only 40 were shown in Table 3) that are valid/invalid associates according to the norms of Nelson et al. and/or the norms of Russell and Jenkins.

Table 4.
Percentages of Responses of WAS model that are Valid/Invalid in Russell & Jenkins (1954) and Nelson et al. (1999) Norms

	Neighbor				
Validity	1	2	3	4	5
valid in Nelson et al.	96	73	61	45	33
valid in Jenkins et al.	96	83	79	69	64
valid in either Nelson et al. or Jenkins et al.	99	86	82	73	66
Invalid in Nelson et al. but valid in Jenkins et al.	3	13	21	28	33

The last row shows that a third of the 5th closest neighbors in WAS are not associates according to the norms of Nelson et al. but that are associates according to the norms of Russell and Jenkins. Therefore, some close neighbors in WAS are valid associates depending on what norms are consulted. However, some close neighbors in WAS are not associates according to either norms. For example, 'angry' is the 2nd neighbor of 'anger' in WAS. These words are obviously related by word form, but they do not to appear as associates in free association tasks because associations of the same word form tend to be edited out by participants. However, because these words have similar associative structures, WAS puts them close together in the vector space.

Also, some close neighbors in WAS are not direct associates of each other but are indirectly associated through a chain of associates. For example, the pairs 'blue-pants' , 'butter-rye', 'comfort-table' are close neighbors in WAS but are not directly associated with each other. It is likely that because WAS is sensitive to indirect relationships in the norms, these word pairs were put close together in WAS because of the indirect associative links through the words 'jeans', 'bread' and 'chair' respectively. In a similar way, 'cottage' and 'cheddar' are close neighbors in WAS because cottage is related (in one meaning of the word) to 'cheese', which is an associate of 'cheddar'.

To summarize, we showed that the output order of words in free association norms is preserved to some degree in WAS: first associates in the norms are likely to be close neighbors in WAS. However, there are some interesting differences between the similarity structure of WAS and the associative strengths of the words in the norms. Words that are not directly associated can be close neighbors in WAS when the words are indirectly associatively related through a chain of associates. Also, although they appear to be exceptions, some words that are directly associated in the norms are not close neighbors in WAS. Because of these differences, WAS is not an exceptionally good model for the task of predicting free association data. However, it is important to realize that WAS was not developed as a model of free association (e.g. Nelson & McEvoy, Dennis, 2000) but rather as a model based on free association.

Semantic/ Associative Similarity Relations

In the priming literature, several authors have tried to make a distinction between semantic and associative word relations in order to tease apart different sources of priming (e.g. Burgess & Lund, 2000; Chiarello, Burgess, Richards & Pollock, 1990; Shelton & Martin, 1992). Burgess and Lund (2000) have argued that the word association norms confound many types of word relations, among them, semantic and associative word relations. Chiarello et al. (1990) give "music" and "art" as examples of words that are semantically related because the words are rated to be members of the same semantic category (Battig & Montague, 1969). However, they claim these words are not associatively related because they are not direct associates of each other (according to the various norm databases that they used). The words "bread" and "mold" were given as examples of words that are not semantically related because they are not rated to be members of the same semantic category but only associatively related (because "bread" is an associate of "mold"). Finally, "cat" and "dog" were given as examples of words that are both semantically and associatively related.

Table 5. Average Similarity between Word Pairs with Different Relations: Semantic, Associative, and Semantic + Associative

Relation	#Pairs	S_{ij}*	k = 10		k = 50		k = 200		k = 400	
Random	200	.000 (.000)	0.34	(0.277)	0.075	(0.178)	0.024	(0.064)	0.017	(0.048)
Word pairs from Chiarello et al. (1990)										
Semantic only	33	.000 (.000)	0.73	(0.255)	0.457	(0.315)	0.268	(0.297)	0.215	(0.321)
Associative only	43	.169 (.153)	0.902	(0.127)	0.83	(0.178)	0.712	(0.262)	0.666	(0.289)
Semantic + Associative	44	.290 (.198)	0.962	(0.053)	0.926	(0.097)	0.879	(0.180)	0.829	(0.209)
Word pairs from Shelton and Martin (1992)										
Semantic only	26	.000 (.000)	0.724	(0.235)	0.448	(0.311)	0.245	(0.291)	0.166	(0.281)
Semantic + Associative	35	.367 (.250)	0.926	(0.088)	0.929	(0.155)	0.874	(0.204)	0.836	(0.227)

Note: standard deviations given between parentheses

*S_{ij} = average forward and backward associative strength = (A_{ij} + A_{ji}) / 2

We agree that the responses in free association norms can be related to the cues in many different ways, but it seems very hard and perhaps counterproductive to classify responses as purely semantic or purely associative[9]. For example, word pairs might not be directly but indirectly associated through a chain of associates. The question then becomes, how much semantic information do the free association norms contain beyond the direct associations? Because WAS is sensitive to the indirect associative relationships between words, we took the various examples of word pairs given by Chiarello et al. (1990) and Shelton and Martin (1992) and computed the WAS similarities between these words for different dimensionalities as shown in Table 5.

In Table 5, the interesting comparison is between the similarities for the semantic only related word pairs[10] (as listed by Chiarello et al., 1990) and 200 random word pairs. The random word pairs were selected to have zero forward and backward associative strength.

It can be observed that the semantic only related word pairs have higher similarity in WAS than the random word pairs. Therefore, even though Chiarello et al. (1990) have tried to create word pairs that were only semantically related, WAS can distinguish between word pairs that were judged to be only semantically related and random word pairs both of which having no direct associations between them. This discrimination is possible because WAS is sensitive to indirect associative relationships between words. The table also shows that for low dimensionalities, there is not as much difference between the similarity of word pairs that supposedly are semantically only and associatively only related. For higher dimensionalities, this difference becomes larger as WAS becomes more sensitive in representing more of the direct associative relationships.

To conclude, it is difficult to distinguish between pure semantic and pure associative relationships. What some researchers previously have considered to be pure semantic word relations, were word pairs that were related in meaning but that were not directly associated with each other. This does not mean, however, that these words were not associatively related because the information in free association norms goes beyond that of direct associative strengths. In fact, the similarity structure of WAS turns out to be sensitive to the similarities that were argued by some researchers to be purely semantic. We will illustrate this point with the following experiment.

Predicting similarity ratings

To further assess the degree to which WAS is capable of capturing semantic similarity, the similarity structure in WAS was compared to semantic similarity ratings collected by Romney et al. (1993). Based on the triads method, they constructed perceived similarity matrices for different sets of 21 words that fell in various semantic categories. Of these categories, we took the 10 matrices for the 10 categories of birds, clothing, fish, fruits, furniture, sports, tools, vegetables, vehicles, and weapons. Between 23 and 28 subjects assessed the similarity of members within each category. Because subjects assessed the similarity for members within categories (e.g. for different birds), we thought it would be a

very challenging task to capture the within category similarity structure (as judged by subjects) with methods such as WAS or LSA.

The correlation was calculated between all observed similarities and predicted similarities using the different methods of WAS, associative strengths and LSA. The result for different dimensionalities is shown in Figure 2, bottom right panel. All correlations were relatively low with the best correlation of 0.36 obtained for WAS based on the SVD of $S^{(1)}$. Interestingly, the semantic spaces of WAS led to higher correlations than LSA. This difference provides further evidence that the similarity structure in word association norms and the similarity structure extracted by WAS from the word association norms does contain many aspects of semantic similarity. Therefore, the similarity structure in word association norms or WAS cannot be simply categorized as 'associative', nor can one conclude that there is no semantic structure.

Predicting Asymmetries in Word Association Norms

In the free association norms, a large number of word associations are asymmetric. For example, the cue "fib" is strongly associated with "lie", but not vice versa. In WAS, the distance between word A and B is by definition equivalent to the distance between words B and A. This characteristic raises the question of how we are able to account for the asymmetries in free associations with WAS. Nosofsky (1991), and Ash and Ebenholtz (1962) have argued that asymmetries in judgments are not necessarily due to asymmetries in the underlying similarity relationships. With free association norms, if the association strength A->B is stronger than B->A, a simple explanation could be that word B has higher word frequency, is more salient and/or, its representation is more available than word A. To check whether word frequency can be used to explain asymmetries in the free association norms of Nelson et al., we took the 500 strongest asymmetric associative word pairs and assigned the more often produced response to B, and the less often produced response to A (in other words: A->B is stronger than B->A). In Figure 3a, the Kucera and Francis (1967) word frequency of A is plotted against the word frequency of B. The diagonal line represents the case where the word frequencies of A and B are equal. It can be seen that most of these word associations fall above the diagonal (463 out the 500, or 92.6%), so word frequency can indeed be used to explain the directionality of word associations.

Krumhansl (1978) has proposed that asymmetries can also arise because of differences in the density of exemplars in parts of the psychological space. For example, if A->B is stronger than B->A, this could be explained because B is placed in a dense region of the space whereas A is placed in a more sparse region of the space. In sampling models, these differences in density might lead to different results. In the previous example, sampling B taking A as a reference would be a more likely event than sampling A taking B as a reference when A has few close neighbors other than B, and B has many close neighbors other than A. We

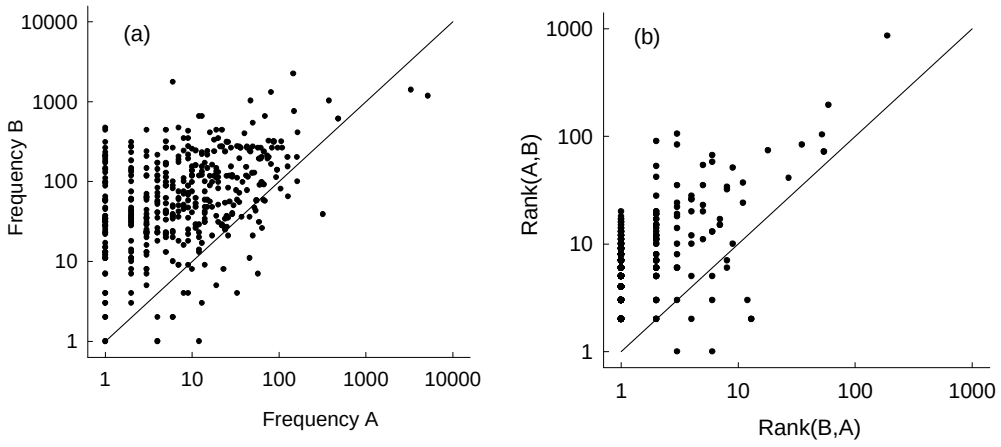

Figure 3. In (a), the Kucera & Francis word frequency is plotted for words A versus words B of word pairs that are strongly associated in the direction A->B but not vice versa. In (b), the plot shows that for word pairs that are strongly associated from A->B but not vice versa, B is often a closer neighbor to A than A is to B in WAS

checked whether this mechanism can explain the directionality of the 500 asymmetric word associations. Instead of calculating density in local regions of WAS, we calculated two numbers for each of the 500 asymmetric word pairs. First, we calculated the rank of the similarity of A-B among the similarity of A to all other words (this will be denoted by rank(B,A)) and among the similarity of B to all other words (this will be denoted by rank(A,B)). For example, if B is the 2nd neighbor of A and A is the 3rd neighbor of B, then rank(B,A)=2 and rank(A,B)=3. In Figure 3b, rank(A,B) is plotted against rank(B,A). The diagonal indicates the cases where the number of neighbors separating A and B is the same for A and B. Most of the cases (486 out of 500, or 97.2%) lie above the diagonal, meaning that in WAS, B is closer to A than A is to B in the sense of number of neighbors separating A and B. This means that Krumhansl's (1978) idea of differences in density can be applied to WAS to explain the directionality of word associations.

Discussion

It has been proposed that various aspects of words can be represented by separate collections of features that code for temporal, spatial, frequency, modality, orthographic, acoustic, and associative aspects of the words (Anisfeld & Knapp, 1968; Bower, 1967; Herriot, 1974; Underwood, 1969; Wickens, 1972). In this research, we have focused on the associative/semantic aspects of words.

By a statistical analysis of a large database of free association norms, the Word Association Space (WAS) was developed. In this space, words that have similar associative structures are placed in similar regions of the space. In the first version of WAS, singular value decomposition was applied on the direct associations between words to place these words in a high dimensional semantic space. In the second version of WAS, the same technique was applied on the direct and indirect associations between words. In the third version of WAS, metric multidimensional scaling was applied on measures for the associative strength related to the shortest associative path between words (similar to the approach in Cooke et al., 1986 and Klauer & Carroll, 1995).

Because the free association norms have been an integral part in predicting episodic memory phenomena (e.g. Cramer, 1968; Deese, 1965; Nelson, Schreiber, & McEvoy, 1992), it was assumed that a semantic space based on free association norms would be an especially useful construct to model memory phenomena. We compared WAS with LSA in predicting the results of several memory tasks: similarity ratings in recognition memory, percentage correct in extra cued recall and intrusion rates in free recall. In all these memory tasks, WAS was a better predictor for performance than LSA. This suggests to us that WAS forms a useful representational basis for memory models that are designed to store and retrieve words as vectors of feature values.

Many memory models assume that the semantic aspects of words can be represented by collections of features abstractly represented by vectors (e.g. Eich, 1982; Murdock, 1982; Pike, 1984; Hintzman, 1984, 1988; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997, 1998). However, in most memory modeling, the vectors representing words are arbitrarily chosen and are not based on or derived by some analysis of the meaning of actual words in our language. We expect that memory models based on these semantic vectors from WAS will be useful for making predictions about the effects of varying semantic similarity in memory experiments for individual words.

For the first two versions of WAS that were based on singular value decomposition, the number of dimensions of the semantic space that led to the best fits with observed performance varied between 200 and 500 dimensions. A similar range in the number of dimensions has shown to be effective in the LSA approach (Landauer & Dumais, 1997). Interestingly, the third version of WAS based on metric multidimensional scaling achieved nearly the same performance but with only 20-40 dimensions. This suggests that the semantic vectors in memory models might involve relatively few feature values to capture the semantic similarity structure for a few thousand words.

We propose that WAS is an approach that augments other existing methods available for placing words in a psychological space. It differs from the LSA and HAL approaches in several ways. LSA and HAL are automatic methods and do not require any extensive data collection of ratings or free associations. With LSA and HAL, tens of thousands of words can be placed in the space, whereas in WAS, the number of words that can be placed depends on the number of words that can be normed. It took Nelson et al. (1999) more than a decade to collect the norms, highlighting the enormous human overhead of the method. Even though a working vocabulary of about 5000 words in WAS is much smaller than the 70,000 word long vocabularies of LSA and HAL, we believe it is large enough for the purpose of modeling performance in variety of memory experiments. An advantage of LSA and HAL is that these approaches have the potential to model the learning process that a language learner goes through. For example, by

feeding the LSA or HAL model successively larger chunks of text, the effect that learning has on the similarity structures of words in LSA or HAL can be simulated. In WAS, it is in principle possible to model a language learning process by collecting free association norms for participants at different stages of the learning process. In practice however, such an approach would not easily be accomplished. To conclude, we think that the WAS, LSA, and HAL approaches to creating semantic spaces are all useful for theoretical and empirical research and that the usefulness of a particular space will depend on the particular task it is applied to.

Author Note

The Word Association Space (WAS) vectors can be downloaded in Matlab format from (www address will be given in future drafts). We would like to thank Tom Landauer and Darrell Laham for kindly providing us with the LSA vectors for the encyclopedia and tasa corpus. Also, we would like to acknowledge Josh Tenenbaum's help in figuring out ways to apply metric multidimensional scaling techniques on the word association norms. Correspondence concerning this article can be addressed to Mark Steyvers, Psychology Department, Stanford University, Stanford, CA 94305-2130. E-mail may be sent to msteyver@psych.stanford.edu.

Notes

1. This method was developed by Osgood, Suci, and Tannenbaum (1957). Words are rated on a set of bipolar rating scales. The bipolar rating scales are semantic scales defined by pairs of polar adjectives (e.g. "good-bad", "altruistic-egotistic", "hot-cold"). Each word that one wants to place in the semantic space is judged on these scales. If numbers are assigned from low to high for the left to right word of a bipolar pair, then the word "dictator" for example, might be judged high on the "good-bad", high on the "altruistic-egotistic" and neutral on the "hot-cold" scale. For each word, the ratings averaged over a large number of subjects define the coordinates of the word in the semantic space. Because semantically similar words are likely to receive similar ratings, they are likely to be located in similar regions of the semantic space. The advantage of the semantic differential method is its simplicity and intuitive appeal. The problem inherent to this approach is the arbitrariness in choosing the set of semantic scales as well as the number of such scales.

2. Because we were unable to obtain the HAL vectors, we could not include HAL in the comparisons in this paper.

3. The result of applying SVD to an asymmetric word association matrix would be two vector spaces: these would capture the commonalities among the cues (rows) and responses (columns) respectively. Analyses showed that neither of these vector spaces captures the variance in the memory tasks reported in this paper as much as a vector space derived at by an SVD on a symmetrized word association matrix.

4. The number of dimensions that can be extracted is constrained by various computational aspects. We were able to extract only the first 500 dimensions in the three scaling solutions of WAS based on 5018 words and 400 dimensions in the scaling solution based on all 5018 words in the free association database.

5. Various alternative scaling procedures such as nonmetric MDS can handle missing values so that not all distances in the matrix would have to be estimated. However, nonmetric MDS cannot be applied to matrices with thousands of rows and columns so we were forced to with the metric MDS procedure that works with large matrices but that does not allow for missing values.

6. In fact, the matrix $S^{(1)}$ has 99.5% zero entries respectively so estimating the distances for these zero entries is nontrivial.

7. The rating were converted to z-scores by subtracting the mean rating from each rating and then dividing by the standard deviation. This was performed for each subject separately and the results were averaged over subjects. This procedure removed some of the idiosyncratic tendencies of subjects to use only part of the rating.

8. The original database has 2272 cue target pairs. We averaged the percentage correct results over identical cue-target pairs that were used in different experiments. This gave 1115 unique pairs.

9. Since responses in word association tasks are by definition all associatively related to the cue, it is not clear how it is possible to separate the responses as semantically and associatively related.

10. Some word pairs in the semantic only conditions that were not directly associated according to various databases of free association norms were actually directly associated using the Nelson et al. (1998) database. These word pairs were excluded from the analysis.

References

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. Journal of Experimental Psychology, 77, 171-179.

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(3), 1-46.

Brainerd, C.J., & Reyna, V.F. (1998). When things that were never experienced are easier to "remember" than things that were. Psychological Science, 9, 484-489.

Brainerd, C.J., Reyna, V.F., & Mojardin, A.H. (1999). Conjoint recognition. Psychological Review, 106, 160-179.

Bousfield, W.A. (1953). The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 49, 229-240.

Bower, G.H. (1967). A multicomponent theory of the memory trace. In K.W. Spence & J.T. Spence (Eds.), The psychology of learning and motivation, Vol 1. New York: Academic Press.

Burgess, C., Livesay, K., and Lund, K. (1998). Explorations in context space: Words, sentences, discourse. Discourse Processes, 25, 211-257.

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich and A.B. Markman (Eds.), Cognitive dynamics: conceptual and representational change in humans and machines. Lawrence Erlbaum.

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

Canas, J. J. (1990). Associative strength effects in the lexical decision task. The Quarterly Journal of Experimental Psychology, 42, 121-145.

Caramazza, A., Hersch, H., & Torgerson, W.S. (1976). Subjective structures and operations in semantic memory. Journal of verbal learning and verbal behavior, 15, 103-117.

Cooke, N.M., Durso, F.T., & Schvaneveldt, R.W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 538-549.

Cramer, P. (1968). Word Association. NY: Academic Press.

Deese, J. (1959a). Influence of inter-item associative strength upon immediate free recall. Psychological Reports, 5, 305-312.

Deese, J. (1959b). On the prediction of occurrences of particular verbal intrusions in immediate recall. Journal of Experimental Psychology, 58, 17-22.

Deese, J. (1962). On the structure of associative meaning. Psychological Review, 69, 161-175.

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

Derweester, 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.

Eich, J.M. (1982). A composite holographic associative recall model. Psychological Review, 89, 627-661.

Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Herriot, P. (1974). Attributes of memory. London: Methuen.

Hintzman, D.L. (1984). Minerva 2: a simulation model of human memory. Behavior Research Methods, Instruments, and Computers, 16, 96-101.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 95, 528-551.

Jenkins, J.J., Mink, W.D., & Russell, W.A. (1958). Associative clustering as a function of verbal association strength. Psychological Reports, 4, 127-136.

Johnson, R.A., & Wichern, D.W. (1998). Applied multivariate statistical analysis. New Jersey, Prentice Hall.

Krumhansl, C.L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. Psychological Review, 85, 445, 463.

Klauer, K.C., & Carroll, J.D. (1995). Network models for scaling proximity data. In, R.D. Luce, M. D'Zmura, D. Hoffman, G.J. Iverson, & A.K. Romney (eds.), Geometric representations of perceptual phenomena. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kucera, H., & Francis, W.N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

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., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, and Computers, 28, 203-208.

Marshall, G.R., & Cofer, C.N. (1963). Associative indices as measures of word relatedness: a summary and comparison of ten methods. Journal of Verbal Learning and Verbal Behavior, 1, 408-421.

McClelland, J.L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. Psychological Review, 105, 724-760.

Morton, J.A. (1970). A functional model for memory. In D.A. Norman (Ed.), Models of human memory. New York: Academic Press.

Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. Psychological Review, 89, 609-626.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G.W. Humphreys (Eds.), Basic processes in reading: Visual word recognition (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nelson, D.L. (2000). The cued recall database. http://www.usf.edu/~nelson/CuedRecallDatabase.

Nelson, D.L., Bennett, D.J., & Leibert, T.W. (1997). One step is not enough: making better use of association norms to predict cued recall. Memory & Cognition, 25, 785-706.

Nelson, D.L., McEvoy, C.L., & Dennis, S. (2000), What is free association and what does it measure? Memory & Cognition, 28, 887-899.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1999). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation.

Nelson, D.L., McKinney, V.M., Gee, N.R., & Janczura, G.A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. Psychological Review, 105, 299-324.

Nelson, D.L., & Schreiber, T.A. (1992). Word concreteness and word structure as independent determinants of recall. Journal of Memory and Language, 31, 237-260.

Nelson, D.L., Schreiber, T.A., & McEvoy, C.L. (1992). Processing implicit and explicit representations. Psychological Review, 99, 322-348.

Nelson, D.L., Xu, J. (1995). Effects of implicit memory on explicit recall: Set size and word frequency effects. Psychological Research, 57, 203-214.

Nelson, D.L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. Psychonomic Bulletin & Review, 7, XXX-XXX.

Nelson, D.L., Zhang, N., & McKinney, V.M. (submitted). The ties that bind what is known to the recognition of what is new.

Norman, D.A., & Rumelhart, D.E. (1970). A system for perception and memory. In D.A. Norman (Ed.), Models of human memory. New York: Academic Press.

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

Palermo, D.S., & Jenkins, J.J. (1964). Word association norms grade school through college. Minneapolis: University of Minnesota Press.

Payne, D.G., Elie, C.J., Blackwell, J.M., & Neuschatz, J.S. (1996). Memory illusions: Recalling, and recollecting events that never occurred. Journal of Memory and Language, 35, 261-285.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. Psychological Review, 91, 281-293.

Rips, L.J., Shoben, E.J., & Smith, E.E. (1973). Semantic distance and the verification of semantic relations. Journal of verbal learning and verbal behavior, 12, 1-20.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: remembering words not presented on lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 803-814.

Romney, A.K., Brewer, D.D., & Batchelder, W.H. (1993). Predicting clustering from semantic structure. Psychological Science, 4, 28-34.

Russell, W.A., & Jenkins, J.J. (1954). The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Tech. Rep. No. 11, Contract NS-ONR-66216, Office of Naval Research and University of Minnesota.

Schacter, D.L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: false recall and recognition in amnesic patients. Journal of Memory & Language, 35, 319-334.

Schwartz, R.M., & Humphreys, M.S. (1973). Similarity judgments and free recall of unrelated words. Journal of Experimental Psychology, 101, 10-15.

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

Schiffman, S.S., Reynolds, M.L., & Young, F.W. (1981). Introduction to multidimensional scaling; theory, methods, and applications. New York, NY, Academic Press.

Shiffrin, R.M., Huber, D.E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 267-287.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. Psychonomic Bulletin & Review, 4, 145-166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.). Rational models of cognition. (pp. 73-95), Oxford, England: Oxford University Press.

Steyvers, M., & Shiffrin, R.R. (to be submitted). Modeling semantic and orthographic similarity effects on memory for individual words.

Torgeson, W.S. (1952). Multidimensional scaling: I. Theory and method. Psychometrica, 17, 401-419.

Underwood, B.J. (1965). False recognition produced by implicit verbal responses. Journal of Experimental Psychology, 70, 122-129.

Underwood, B.J. (1969). Attributes of memory, Psychological Review, 76, 559-573.

Wickens, D.D. (1972). Characteristics of word encoding. In A.W. Melton & E. Martin (Eds.), Coding processes in human memory. Washington, D.C.: V.H. Winston, pp. 191-215.