Predicting Word-Naming and Lexical Decision Times from a Semantic Space Model

Brendan T. Johns (johns4@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. Tenth St.
Bloomington, In 47405 USA

Michael N. Jones (jonesmn@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. Tenth St.
Bloomington, In 47405 USA

Abstract

We propose a method to derive predictions for single-word retrieval times from a semantic space model trained on text corpora. In Experiment 1 we present a large corpus analysis demonstrating that it is the number of unique semantic contexts a word appears in across language, rather than simply the number of contexts or the frequency of the word, that is the most salient predictor of lexical decision and naming times. In Experiment 2, we develop a co-occurrence learning model that weights new contextual uses of a word based on fit to what currently exists in the word's memory representation, and demonstrate this model's superiority in fitting the human data compared to models built using information about the word's frequency or number of contexts. Finally, in Experiment 3 we find that building lexical representations using semantic distinctiveness naturally produces a better-organized semantic space to make predictions for semantic similarity between words.

Keywords: Co-occurrence model; Lexical-decision; LSA; Contextual distinctiveness

Introduction

The last decade has seen remarkable progress with co-occurrence models of lexical semantics (e.g., Lund & Burgess, 1996; Landauer & Dumais, 1997). These models learn semantic representations for words by observing lexical co-occurrence patterns across a large text corpus, typically representing the words in a high-dimensional semantic space. This approach provides both an account of the semantic representation for words and an account of the learning mechanisms humans use to build and organize semantic memory. Co-occurrence models have seen considerable success at accounting for data in a wide variety of semantic tasks, including TOEFL synonyms (Landauer & Dumais, 1997), semantic similarity ratings and exemplar categorization (Jones & Mewhort, 2007), and free association norms (Griffiths, Steyvers, & Tenenbaum, 2007).

To date, all applications of co-occurrence models have been to semantic similarity *between* two words or two documents. The standard prediction of semantic similarity in these models is some measure of the angle between two vectors. However, co-occurrence models should, in theory, contain sufficient information in the magnitude of their representations to make predictions about single word retrieval as well.

Lexical decision time (LDT) and word naming time (NT) are both important variables that offer insight into the organization of semantic memory. In a lexical decision task, a letter string is presented and the participant provides a speeded response of whether the string is a word or not. In a naming task, the participant's task is to name the presented word aloud as quickly as possible. Both measures produce an index of a word's identification latency. Orthographic and phonological factors are certainly large components of both LDT and NT, but semantics plays a significant role as well, and co-occurrence models have yet to be extended to predicting reaction time variance for these single-word identification tasks.

Modeling of retrieval times is usually done by looking for the best environmental correlates of LDT and NT (Adelman & Brown, 2008). Some of the most influential models of retrieval times are based upon word frequency. Word frequency (WF) has been used to drive many different types of models, including serial-searched rank frequency models (Murray & Forster, 2004), threshold activation models (Coltheart, et al., 2001), and connectionist models (Seidenberg & McClelland, 1989).

However, recent evidence suggests that word frequency may not drive retrieval times but, rather, the causal factor is a word's contextual diversity (Adelman, Brown, & Quesada, 2006; Adelman & Brown, 2008). Contextual diversity (CD) is the number of different contexts that a word appears in, and is based on the rational analysis of memory (Anderson & Milson, 1989), particularly the *principle of likely need* (PLN). PLN states that the more unique contexts a word appears in, the more likely the word will be needed in any future context. Hence, a word with a high CD should be faster to retrieve under this principle.

A word's CD value is typically computed by simply counting the number of different documents in which it appears across a text corpus. This measure has been shown to be a better predictor of LDT and NT than WF (Adelman, et al., 2006). However, operationalizing CD as the number of documents in which a word occurs may not be a fair instantiation of PLN. A word that appears in many documents may have a high WF, but it should have a low CD if those documents are highly redundant, as is the case with words that belong to a popular discourse topic for which many documents exist. It is the number of different contexts *and* the uniqueness of contexts that determines a word's likely need. This calls for a measure of CD that considers the semantic uniqueness of documents that a word appears in. Based on PLN, it is reasonable to assume that if a word appears in a context it has never before occurred in,

then the context should be more important to its CD count. A greater number of unique contexts should yield a higher CD value than an equal number of redundant contexts. This interpretation of PLN has empirical support from experiments demonstrating that benefit from repeated exposure to an item is strongest if the context changes as well (Glenberg, 1979; Verkoeijen, Rikers, & Schmidt, 2004).

There is no principled reason why a co-occurrence model could not compute a WF or document count to make word-specific LDT or NT predictions. Most co-occurrence models begin with a word-by-document matrix, which contains the requisite information in the magnitude (sum) of a word's frequency distribution over documents. In Experiment 1, we conduct a large corpus-based analysis to demonstrate that number of unique contexts is a more important factor than a simple document count or WF in predicting LDT and NT. In Experiment 2 we develop a co-occurrence model that learns from semantic distinctiveness and makes retrieval predictions, and in Experiment 3 we demonstrate that in addition to giving a better fit to the LDT and NT data than frequency or document count models, our model based on semantic uniqueness naturally produces better predictions of semantic similarity between words as well.

Experiment 1: Corpus Analysis

Semantic Distinctiveness

To examine the influence of semantic distinctiveness, it is necessary to create a measure of the coherence of documents in which a word appears. Though there are many existing models of semantic representation (e.g., HAL or LSA), we did not want to approach the problem from a specific theoretical orientation. Instead the measure that we use to assess the dissimilarity between two documents is based on the proportion of words that two documents have in common, or:

$$Dissim(doc1, doc2) = 1 - \frac{|doc1 \cap doc2|}{|\min(doc1, doc2)|} \quad (1)$$

That is, document similarity is the intersection of the two sets of words, divided by the size of the smaller document. This gives the proportion of word overlap between two documents. Function words (e.g. *the, is, of*, etc…) were filtered out of the set of words, so they do not impact the similarity rating. Document dissimilarity is then just 1-similarity. We then define a word's *semantic distinctiveness* (SD) as the mean dissimilarity of the set of documents that contain it:

$$SD_{word} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{i} Dissim(doc_i, doc_j)}{n + (n^2 - n)/2} \quad (2)$$

Where *n* is the number of documents that a word appears in. Equation (2) is the average dissimilarity among all documents that the word appears in, and this SD value

signals how distinct the documents that a word occurs in are from each other. A word with a high SD value tends to occur in documents that have a low amount of word overlap (it is more contextually distinct), and a word with a low SD value tends to occur in documents that have a high amount of word overlap (it is less distinct).

However, as Adelman, et al. (2006) showed, the number of different contexts that a word appears in is a highly important predictor of LDT and NT. These SD values do not take this important source of information into account.

To explore whether counting low similarity contexts as being more important yields a better CD count, the weights given to the value of a context were modified with increasing specificity. This was done by creating increasingly more specific rules, based on the computed SD values, to create the context value. The first iteration will be one rule – if the dissimilarity between two documents is greater than 0.0 (covering 100% of the SD data), then 1 is added to the word's count value (note that this is the same as the standard document count that weights each document equally). On the next iteration, there will be two rules – if the dissimilarity between two documents is greater than the median of the computed SD values, then the count gets two added to it, otherwise (i.e. if it is less than the median) then the count gets 1 added to it. Then on the third iteration, the rules would increase in resolution:

If dissim(docx, docy) < SD_33_percentile => count +=1
If dissim(docx, docy) < SD_66_percentile => count +=2
If dissim(docx, docy) < SD_100_percentile=> count +=3

This is done up to 10 rules (so the data would be split into tenths). By this method, we create a document count in which documents that have more unique contextual uses of the word (compared with the other documents that the word appears in) are weighted more strongly than documents that have more common contextual usages of the word, consistent with PLN.

Method

Our analyses are based on three corpora: 1) TASA (from Touchstone Applied Sciences Associates), 2) a Wikipedia (WIKI) corpus, and 3) a New York Times (NYT) corpus. The TASA corpus was composed of 10,500 documents, with each document having a mean length of 289 words. The Wikipedia corpus was composed of 9,755 documents, with a mean document length of 391 words. The New York Times corpus is composed of 9,100 documents with a mean length of 250 words, drawn from the New York Times during the year of 1994. These are smaller versions of the full corpora, and the reduced size was necessary for practical reasons of computation time: The SD counts took, on average, 120 hours in parallel across 3 Sun Sparc IV+ CPUs for each corpus. LDT and NTs were attained from the English Lexicon Project (Balota, et al., 2002). SD values were computed for 17,984, 22,673, and 14,609 words, for the TASA, WIKI, and NYT corpora, respectively.

Table 2. Naming Time

Results

Figure 1 shows the increase in R^2 for LDT (top panel) and NT (bottom panel) as predicted by the SD weighted context count over DC and WF. These figures show a large increase for the weighted SD counts over both WF and DC in predicting variance for word identification measures. As these figures illustrate, giving greater weight to a context that is more distinct given a word's history of contexts can produce a better fit to the latency measures. In our subsequent analyses, we will use a split of seven quantiles, since after this point there does not appear to be a significant increase in variance predicted for any of the corpora.
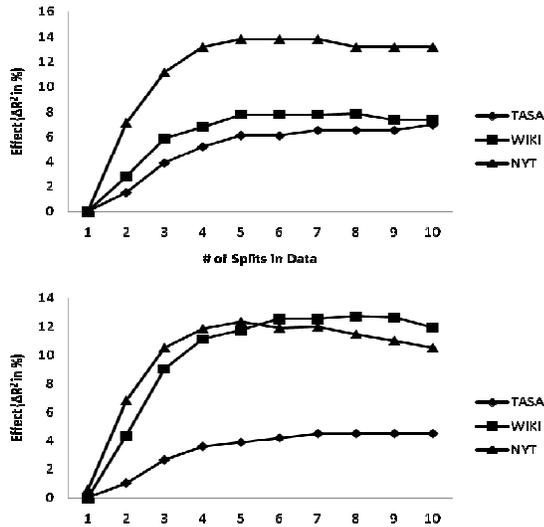

Figure 1. Increase in R^2 over WF and document count predicted by the weighted SD count for LDT (top panel) and NT (bottom panel).

Adelman et al. (2006) found a small but reliable increase in variance predicted for LDT and NT by document count over WF (using log or power transforms of both variables). We conducted a similar regression analysis using our SD_Count, WF, and document count (DC) to predict the

Table 1. Lexical Decision

Analysis	Effect (ΔR^2 in %)		
	TASA	WIKI	NYT
Log_SD (After WF)	5.501	6.417	6.282
Log_CD (After WF)	2.341	1.675	0.0 *n.s.*
Log_ SD (After CD)	3.87	6.807	11.557
Log_WF (After CD)	0.0 *n.s.*	0.382	1.123
Log_CD (After SD)	0.645	2.094	5.025
Log_WF (After SD)	0.0 *n.s.*	0.0 *n.s.*	0.0 *n.s.*
Log_SD(After CD,WF)	4.487	7.731	11.881
Log_WF(After SD, CD)	1.282	1.03	1.485
Log_CD(After SD, WF)	0.641	3.108	5.445

Table 2. Naming Time

Analysis	Effect (ΔR^2 in %)		
	TASA	WIKI	NYT
Log_SD (After WF)	8.49	9.016	7.751
Log_CD (After WF)	3.98	2.654	0.0 *n.s.*
Log_ SD (After CD)	6.511	11.718	13.235
Log_WF (After CD)	0.217	0.0 *n.s.*	0.847
Log_CD (After SD)	0.471	5.468	6.617
Log_WF (After SD)	1.86	0.819	1.55
Log_SD(After CD,WF)	6.511	12.403	13.868
Log_WF(After SD, CD)	1.86	0.775	1.459
Log_CD(After SD, WF)	0.465	5.833	6.569

behavioral measures. Tables 1 and 2 show for LDT and NT, respectively, the unique variance predicted by each measure while the other measures are systematically partialled out. The results in Tables 1 and 2 are similar to those attained with power and rank transformations. The SD_Count variable gives a better prediction of the latencies for every analysis, and wipes out the effect of WF just as well as the document count variable does

Discussion

The results of our corpus analysis clearly suggest that in order to make an accurate contextual diversity measurement one has to take into account the uniqueness of the contexts that a word appears in. Considering the semantic distinctiveness of the contexts that a word appears in, we were able to create a count that is significantly better than one that weights all documents as being equally unique.

Next, we propose a simple process model to create a term-by-document matrix that incrementally weights new contexts for words by considering how distinct a document is at time *t* relative to the word's current lexical representation (which represents the knowledge representation from documents *1...t-1*). We then show how the representation magnitude can be used to predict LDT and NT, and how the weighted input matrix naturally produces a better semantic space as a byproduct of this implementation of PLN.

Experiment 2: Learning Model

A Contextual Relatedness Episodic Activation Model

We next wanted to create a co-occurrence model that can learn semantic distinctiveness and compare predictions on LDT and NT to models that do not. In order to capture the results of the corpus analysis, a contextual relatedness episodic activation memory (CREAM) model was created. Like in other co-occurrence learning models, a word-by-document matrix is built up to create a word's representation. The modification that this model makes is the type of information that is added into the word-by-document matrix: instead of raw frequency or occurrence,

we will add a semantic distinctiveness value. The first step in computing this SD value is to create a 'context' or 'document' vector, which we will simply call a composite context vector (CCV). Simply, for each word that occurs in a document ($W_1,...,W_N$) we add each word's vector into a composite vector representing the meaning of the document. Formally, this is:

$$CCV = \sum_{i=1}^{N} T_i \qquad (3)$$

Where N is the set of words in the document, and T_i is the memory trace corresponding to word i. The next step is to compute a similarity value (given by a vector cosine) between each word that occurs in the context and the context vector. Then this similarity value is transferred through an exponential probability density function, and the resulting value is entered into the new context slot in memory:

$$SD = \mathrm{e}^{-\lambda * cosine} \qquad (4)$$

Where λ is a fixed parameter with a small positive value. This exponential function has the effect of transforming a low similarity value into a large SD value and a high similarity value into a small SD value, as well as smoothing the added value of uniqueness. The parameter λ is much like the weighting scheme that we employed in the corpus analysis. With a small λ (<1), the transformation from a high to low value is almost linear (e.g. 0.9 to 0.1). However, as we increase λ, it accentuates the difference in the value of high vs. low similarity contexts. A document count model can be considered to be nested within this model, with a λ set at 0. As in the corpus analysis, a context with a high SD value means that the document is more distinct compared with the other contexts that a word has appeared in. This gives greater salience to low similarity contexts, in terms of the word's magnitude, than high similarity contexts.

When a word is first seen its context vector will be empty, hence, the similarity between memory and the current context will be 0.0. Therefore, the SD value will always be 1.0 for the first document, and it will be encoded at maximal strength. The second time a word is experienced, the similarity of this context is compared to the word's current lexical representation (which only contains the first context so far). If this is a repetition of the first document, the new context will be encoded at minimal strength. If, however, it is a context that is unique from the first, the new context will be encoded at maximal strength.

In this fashion the word-by-document matrix has columns added to it each time a new document is learned, with the encoding strength for a document (for a particular word) dependent on the lack of fit between what has been learned and what is being experienced. In a sense, the attention weight given to a new context is dependent on how unique the context is relative to the word's current memory representation. That is, creating a word's representation in memory is a dynamic interaction between what is in memory and what is in the environment.

Method

To predict LDT and NT, the magnitude of a word is computed by summing all of the entries in the word's context vector. This magnitude is used as a direct predictor of retrieval times. To judge this model's ability to predict both LDT and NT, a model comparison was undertaken. CREAM was compared against a WF model and a document count (DC) model. In the CREAM model, the λ parameter was fixed at 5.5. In the WF model the frequency that a word occurs in a document is the entry into the word-by-document matrix. In the DC model a 1.0 is entered into the matrix if the word occurs in that document. For all three models, vector magnitude is used to predict latencies; the only difference is how the matrix is built.

This comparison was conducted for the same three corpora as specified in the corpus analysis. However, the models were trained on the full versions of each corpus: 36,700 documents from TASA, with an average length of 121 per document, and 40,000 documents from the Wikipedia corpus, with an average document length of 279. The New York Times corpus was the same as specified in the corpus analysis. LDT and NT data were again attained from the Elexicon database (Balota et al., 2000). In the analysis, latencies from 29,799, 35,518, and 20,744 words were used for the TASA, WIKI, and NYT corpora, respectively.

Table 3. Lexical Decision

	Effect (ΔR^2 in %)		
	TASA	WIKI	NYT
Log_CREAM (After WF)	3.048	1.81	5.461
Log_DC (After WF)	1.274	0.786	0.0 *n.s.*
Log_ CREAM (After DC)	2.346	0.849	6.901
Log_WF (After DC)	0.0 *n.s.*	0.364	1.07
Log_DC (After CREAM)	0.511	0.141	0.462
Log_WF (After CREAM)	0.0 *n.s.*	0.704	0.0 *n.s.*
Log_CREAM(After DC,WF)	3.118	1.175	7.348
Log_DC(After CREAM, WF)	1.327	0.149	2.001
Log_WF(After CREAM, DC)	0.816	0.7	1.549

Table 4. Naming Time

	Effect (ΔR^2 in %)		
	TASA	WIKI	NYT
Log_CREAM (After WF)	5.811	3.323	6.568
Log_DC (After WF)	2.904	2.01	0.0 *n.s.*
Log_ CREAM (After DC)	4.984	1.213	7.791
Log_WF (After DC)	0.119	0.0 *n.s.*	0.75
Log_DC (After CREAM)	2.062	0.0 *n.s.*	0.72
Log_WF (After CREAM)	0.386	0.132	0.0 *n.s.*
Log_CREAM(After DC,WF)	5.361	1.336	8.243
Log_DC(After CREAM, WF)	2.163	0.0 *n.s.*	1.868
Log_WF(After Sem, DC)	0.485	0.117	1.197

Results

Table 3 shows the increase in R^2 for the various models fit to LDT, while controlling for the other models' magnitudes. Table 4 shows the same data for NT. As the tables show, the

CREAM model performs significantly better then both the WF and DC model. As was suggested by the corpus analysis, this comparison demonstrates that by weighting the context for how unique it is to other contexts that the word has appeared in, we can create a better contextual diversity count that produces closer correspondence to the behavioral measures.

Experiment 3: Semantic Space

Since the word-by-document matrix is used to derive semantic representations in many co-occurrence models, we wanted to test the semantic organization constructed with the SD-weighted learning mechanism against those based on raw frequency or document count. To this end, we adapted the Constructed Semantics model of Kwantes (2005) to derive a semantic space from our matrices. In Kwantes' model a word-by-document matrix is constructed, with each entry in the matrix being the frequency that a word occurs in a specific context. This is the same matrix that both LSA (Landauer & Dumais, 1996) and the Topics Model (Griffiths et al., 2007) begin with. However, instead of using a matrix reduction technique (LSA uses singular value decomposition), the Constructed Semantics model retrieves a word's semantic representation using a retrieval process borrowed from a well-known global memory model – MINERVA 2 (Hintzman, 1986), with a few minor adjustments. By Kwantes' account, the raw instances of a word's context occurrences are stored in memory, and a word's meaning is constructed by the retrieval process, much in the same way MINERVA 2 explains schema abstraction tasks as retrieval from episodic memory (Hintzman, 1986). The retrieval process works by receiving a probe P (which is a word's context vector), and creating a composite vector from memory:

$$C_i = \sum_{j=1}^{M} T_{ij} * A_j \quad (5)$$

where M is the total number of traces in memory (number of words in the word-by-document matrix), T is the current trace, and A is activation of memory trace T.

The activation value is simply the vector cosine (normalized dot product) between the probe and a memory trace. One modification of Kwantes' (2005) model is that if the similarity value is less than a criterion, then this vector is not added into the composite vector.

We use Kwantes' (2005) model here because it is easy to understand memory as the word-by-document matrix. Typically this matrix is decomposed to determine the latent factors that underlie the maximum amount of variance in the original matrix, and semantic similarities emerge as a byproduct of this dimensional reduction. Semantic similarity between words is determined by a measure of the angle between their vectors in this reduced space; for Kwantes' model, this would be the cosine of the vectors retrieved from memory for the two words, and this process

produces semantic structure that is very similar to models such as LSA.

As with other co-occurrence models, Kwantes' (2005) model is able to account for the knowledge representation of words. It is possible that by building the co-occurrence matrix using semantic uniqueness we improved predictions of LDT and NT at the cost of semantic organization. Figure 3 displays a two-dimensional scaling solution for a subset of our space, illustrating that semantically similar words seem to be clustering in intuitive clusters, as is expected. In this figure, we also display a color coding of the magnitude of word vectors (likely need) in addition to semantic organization. From this figure, it does not appear that learning magnitudes based on semantic distinctiveness has sacrificed semantic organization.
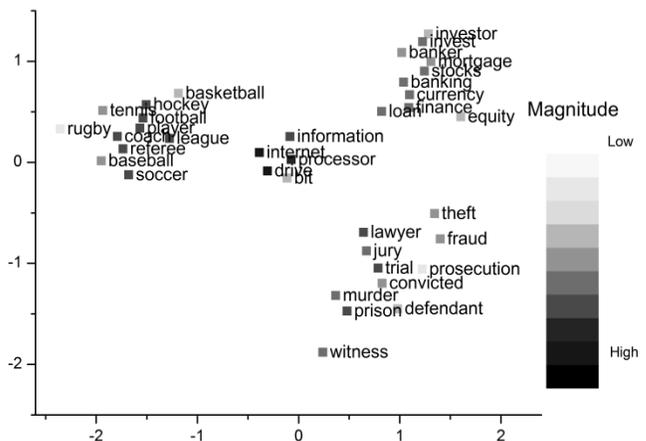

Figure 2. MDS plot depicting organization in a subset of the semantic space. The location of a point indicates its semantic position, and the color indicates vector magnitude (used to predict LDT and NT).

To get a quantitative test of semantic organization we compared the CREAM, WF, and DC models' predictions of semantic distance in WordNet for pairs of words. We used the well known Jiang-Conrath (JCN) semantic distance from WordNet that has been argued gives the best predictions of human judgments of semantic similarities (Maki, McKinley, & Thompson, 2004). To test the models, 730 word pairs from Maki et al.'s (2004) database were created for each model by retrieving their semantic representations from memory, as in equation (5). The retrieval process is the same for all models, only the way in which the word-by-document matrix is constructed differs. The results were also compared against the results of two established models of semantic memory: LSA (Landauer & Dumais, 1997) and BEAGLE (Jones & Mewhort, 2007). The correlation matrix of this analysis is given in Table 5.

Table 5. WordNet Similarity Correlations

Variable	1.	2.	3.	4.	5.	6.
1.JCN	-	-.207	-.332	-.237	-.273	-.287
2. LSA		-	.359	.579	.445	.281
3. BEAGLE			-	.483	.540	.617
4. WF				-	.836	.645
5. DC					-	.857
6. CREAM						-

The CREAM model, based on semantic distinctiveness co-occurrence counts, produced predictions that were significantly closer to the WordNet similarity values than the other models, with the exception of the BEAGLE model. The size of this comparison was limited to 730 words due to the enormous computations required to create semantic representations using Kwantes' (2005) model. However, it cleanly demonstrates that a model based on semantic distinctiveness can produce better predictions of LDT and NT from its vector magnitudes, and, as a byproduct, uniqueness might also produce a better organized semantic space.

This comparison provides converging evidence for the importance of uniqueness of a context (relative to memory) when building a model of contextual co-occurrences, and it proposes a simple process mechanism to build these representations that is based on likely need. The resulting space seems to produce a better approximation of semantic similarity, and the magnitude of a word's vector can then be used to make direct predictions about LDT and NT.

General Discussion

The results of Experiment 1 are consistent with the PLN theme advanced by Adelman et al. (2006). However, they clearly demonstrate that the semantic uniqueness of a context is an important factor to weight when creating a contextual diversity measure. Experiment 2 explored this theme with confirmatory modeling by building a co-occurrence representation in which the novelty of information being learned was contrasted with existing memory structure. It was demonstrated that by doing this the model produces better estimates of LDT and NT than one based simply on frequency or document occurrence. Of interest were the results of Experiment 3: building a semantic space from the superior uniqueness matrix actually seems to produce better semantic organization, a free lunch we are pleased with, but were not explicitly trying to create.

We believe that using the magnitude of a word's vector to predict LDT and NT is a step in the right direction for a unified class of co-occurrence learning models that have already proven success at accounting for a wealth of data. However, it is important to note that our current account is largely exploratory, and currently lacks a sufficient process mechanism to explain the host of strategic effects that are seen as a function of SOA in lexical decision and naming tasks.

References

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity

Adelman, J. S., Brown, G. D. A, Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision time. *Psychological Science*, 17, 814-823.

Anderson, J.R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703-719.

Balota, D.A., Cortese, M.J., Hutchinson, K.A., Neely, Nelson, D., Simpson, G.B., & Treiman, R. (2000). The English Lexicon Project. Retrieved September 30, 2007, from http://elexicon.wustl.edu/.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.

Glenberg, A.M. (1979). Component-levels theory of the effects of spacing of recall and recognition. Memory & Cognition, 7, 95-112.

Griffiths, T.L., Steyvers, M., & Tenenbaum, J.B. (2007). Topics in Semantic Representation. *Psychological Review*, 114, 211-244.

Hintzman, D.L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review,* 93, 411-428.

Jones, M.N., & Mewhort, D.J.K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

Kwantes, P.J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, 12, 703-710.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review,* 104, 211-240.

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

Maki, W.s., McKinley, L.N., & Thompson, A.G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers*, 36, 421-431.

Murray, W.S., & Forster, K.I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111, 721-756.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

Verhoeijen, P.P.J.L., Rikers, R.M.P.J., & Schmidt, H.G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 796-800.