Rethinking context availability for concrete and abstract words: a corpus study

Robert Liebscher and David Groppe
Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0515
{rliebsch|dgroppe}@cogsci.ucsd.edu

Abstract

Over the past five decades, psycholinguists have uncovered robust differences in the processing of concrete and abstract words. One of these findings is that it is easier for people to generate possible contexts for concrete words than for abstract words; that is, concrete words seem to have higher "context availability" (CA). It is not clear why this difference exists, but some have suggested that concrete words may be used in a smaller variety of semantic contexts. In this paper, we review the relevant psycholinguistic literature and report on a previous corpus-based attempt to investigate this property of abstract and concrete words. We then extend the current methods by introducing an information theoretic measure that we use to test the validity of CA. The result runs counter to current thinking in psycholinguistics and suggests a rethinking of context availability.

1. Introduction

In general, people are better at remembering and faster at processing concrete words than abstract words. Although there are exceptions, multiple studies have found this to be true (for reviews see: Paivio 1991; Schwanenflugel 1991). For example, concrete words are identified faster in a lexical decision task (James 1975; Kounios and Holcomb 1994; Kroll and Merves 1986; Schwanenflugel and Shoben 1983) and recalled better in free recall (Nittono 2002; Paivio 1986).

It is not yet clear why these "concreteness effects" exist. One of the most influential theories is the *Context Availability Model* (CAM) of comprehension (Bransford and McCarrell 1974; Kieras 1978; Schwanenflugel and Shoben 1983). According to CAM, linguistic comprehension depends critically on supportive contextual information either from the comprehender's knowledge base or from the external linguistic context. For example, when an experimental subject is asked to judge if the isolated string *apple* is a word or non-word (the lexical decision task), CAM posits that the subject's decision is aided by the generation of a meaningful context for the word (e.g. a person eating an apple). A subject can receive the same aid from external contextual information such as a sentence or a visual scene. As contextual information for a word increases, comprehension of and memory for the word also increase.

CAM can explain effects of concreteness on lexical processing because it is subjectively easier for people to generate contexts for concrete words than for abstract words. Several studies have found that subjects' ratings of the difficulty of generating a context for a word are highly correlated with their judgments of the word's concreteness (Altarriba, Bauer and Benvenuto 1999; Schwanenflugel and Shoben 1983).

CAM is an appealing explanation for concreteness effects because it readily explains why differences between processing concrete and abstract words tend to disappear when the words are placed in a meaningful semantic context. (Schwanenflugel and Shoben 1983; Schwanenflugel, Harnishfeger, and Stowe 1988; Schwanenflugel and Stowe 1989). Furthermore, some studies have found that abstract and concrete words with identical context availability ratings are processed identically in lexical decision and memory tasks (Akin 1989; Akin and Schwanenflugel 1990; Schwanenflugel, Harnishfeger, and Stowe 1988). Also, context availability has been found to be a slightly better predictor of reaction time in a lexical decision task than concreteness (Schwanenflugel and Shoben 1983).

Despite this success, the explanatory power of CAM is limited by the fact that it is not known why concrete words are judged to have a greater context availability than abstract words. This theoretical shortcoming limits our understanding of the mechanisms of context generation. Moreover, it hinders our ability to resolve a major debate in the psychology of language meaning; namely, whether concreteness effects are due to the ways in which words are used in language or whether they reflect qualitative differences in semantic content owing to the physical features of concrete concepts (Funnell 2000; Schwanenflugel 1991).

Schwanenflugel (1991: 243) offers two hypotheses concerning the higher context availability of concrete words. The first hypothesis simply states that concrete words occur more frequently than abstract words. The second hypothesis posits that abstract words appear in a greater *variety* of semantic contexts than concrete words; or, equivalently, that concrete words have stronger *contextual constraints*. The latter hypothesis receives some support from Galbraith and Underwood (1973), who found that subjects judged the variety of possible semantic contexts to be greater, on average, for abstract words.

Previous corpus-based work and the present study

Human intuition and behavioural experiments can only tell us so much, and it is with this in mind that we turn to corpus linguistics to investigate differences between abstract and concrete words. Surprisingly, few corpus-based studies in this area have been undertaken (cf. Audet and Burgess 1999; Wiemer-Hastings and Graesser 2000). We review one here, as it is relevant to our work.

Audet and Burgess (1999) tested both of Schwanenflugel's hypotheses by analyzing how a set of concrete and abstract words behaved in a 320 million token corpus of Usenet text. Contrary to the first hypothesis, they found that abstract words had a significantly greater frequency than concrete words. To test the second hypothesis, they measured the *context density* of the two sets of words. A word's context density is simply the percentage of non-zero elements in its co-occurrence vector (see Section 2 for details). The more unique words with which a given word co-occurs in a surrounding window of text, the greater its context density and presumably the greater its range of semantic contexts. Audet and Burgess found that their abstract words had a higher context density than their concrete words and interpreted this as support for Schwanenflugel's second hypothesis.

As we explain in more detail Sections 2 and 3, a problem with the Audet and Burgess result is that context density is potentially confounded by frequency and is a coarse measure of contextual constraints. The purpose of the current study is to retest Schwanenflugel's second hypothesis and to evaluate effects of frequency on the previous Audet and Burgess results. To do so we analyze the contextual constraints on abstract and concrete words in Usenet text using context density and a more principled measure, the entropy of a word's co-occurrence distribution.

2. Methods

Context density

To determine the relative number of contexts in which a given word occurs, Audet and Burgess (1999) devised the *context density* measure. As part of the HAL model (Lund and Burgess 1996), they constructed co-occurrence vectors for every word[1] in a corpus of Usenet text using a ten token weighted window on either side of the word in question. The result is a matrix of 70 000 by 140 000 elements (tokens occurring in the left window and right window are stored separately).

Context density is defined as the proportion of non-zero elements in a word's co-occurrence vector. Using this measure, abstract words were found to have a significantly greater context density than concrete words. Audet and Burgess contend that this finding "suggests that abstract word representations might have more diffuse connections to associated contextual information" (40).

However, inference from the context density measure is problematic for two reasons. While context density should reflect the range of words a given word co-occurs with, it will also reflect the frequency of that word. Each time a word appears in a corpus, there is a chance for the word to co-occur with a word with which it has not previously co-occurred. Thus, the finding that the abstract words used by Audet and Burgess have a higher context density could simply be due to their greater frequency.

Even if one controls for differences in frequency, context density remains problematic as it only reflects the presence of an associative relationship between words, and not the strength of that association. As Audet and Burgess (1999: 40) note:

[1] For simplicity, we use *word* here because we are only interested in the properties of abstract and concrete words, as were Audet and Burgess (1999). The HAL model, however, has been used to investigate other features of language, and as such there are vectors for more than just words; for example, numbers and punctuation symbols are also included.

> We are aware that the *diversity* of contexts in which a word appears is not reflected only in the raw number of contexts for that word. Contextual diversity is also a function of the differences between the number of occurrences of a word in each of the different contexts in which that word was experienced.... Put more simply, two words might occur in roughly the same number of contexts, but vary in the overall pattern of occurrences across these contexts, thus leading to quantitatively and qualitatively different representations.

In the next section we propose a better measure of contextual constraint, the information theoretic measure of entropy, that does not suffer from these shortcomings.

Entropy as a better measure of contextual constraint

We believe that the information theoretic measure of entropy (see, for example, Cover and Thomas 1991), when applied to a word's co-occurrence vector, is an appropriate measure of context *diversity*. From the co-occurrence vector of an individual word w_i, it is calculated by first finding the probability $P(w_j|w_i)$ that w_j will be found in the context window[2] of word w_i:

$$P(w_j \mid w_i) = \frac{C(w_j, w_i)}{\sum_j C_{i,j}}$$

where $C(w_i,w_j)$ is the number of times w_i co-occurs with w_j and the denominator is simply the sum of w_i's vector. The entropy $H(w_i)$ is then:

$$H(w_i) = -\sum_j P(w_j \mid w_i) \cdot \log_2 (P(w_j \mid w_i))$$

The entropy measure reflects not only the co-occurrence of words, but also the relative strength of co-occurrence relationships. A word that co-occurs with n other words will have a relatively high entropy if its distribution of occurrence is fairly uniform over n, and will have a smaller entropy if its distribution of occurrence is concentrated on a subset k of those n words. As k approaches 1, entropy falls to a minimal value of 0; its context is maximally constrained. A maximal entropy is achieved if a word co-occurs with equal frequency with every word in the corpus; its context is minimally constrained. Function words have a very high entropy, whereas words like *diagonalise*, from the lexically limited domain of linear algebra, should have a very low entropy.

Corpus and target word lists

Our corpus contains approximately 230 million tokens of Usenet (Google Groups) text from a variety of electronic discussion forums, including scientific discourse, sports discussions, job postings, and many more. The posts come from thousands of different authors, and are all contemporaneous. Appropriate measures were taken to capture prose while eliminating spurious text such as message headers and signature lines.

We created two different lists of abstract and concrete words for our study. Our goal was to control for factors that might be influencing our measures in ways that are irrelevant to concreteness effects. To this end, we attempted to control for word frequency and concreteness ratings.

The AB list contains words that are known to produce concreteness effects under experimental conditions. Audet and Burgess (1999) constructed a list of 248 such words drawn from the psycholinguistic studies of Bleasdale (1987) and Chiarello, Senehi, and Nuding (1987). From this list, we chose 45 abstract and 45 concrete words that met a minimum frequency threshold of 10 000 uses in our 230 million token corpus, and were assigned concreteness ratings below 425 (for abstract words) and above 550 (for concrete words) on a 100-700 scale, 700 being most concrete. Ratings were obtained from the MRC2 psycholinguistic database (Wilson 1988)[3].

Although the AB list consists mostly of words commonly used as nouns, it also contains some adjectives and verbs. This, coupled with the large variability in concreteness ratings (see Table 1), led us to produce the LG list, which consists entirely of words that are used commonly as nouns and are more prototypically concrete and abstract. It

[2] We used an unweighted window of five words on either side because of the nature of our task. HAL uses a linear weighted window, with words closest to the target word being "counted" more than distant words within the context window. This is useful for elucidating grammatical and syntactic properties, but is not relevant to our task. See Levy and Bullinaria (2001) for details on the properties of different weighting schemes and window sizes.

[3] Concreteness ratings for ten of the abstract words in the **AB** list were not available in the MRC2 database. Ratings for nine of the words were taken from Altariba, Bauer, and Benvenuto (1999) and Nelson, McEvoy, and Schreiber (1998). A rating for the word *current* was not found in any of these sources.

contains 45 abstract and 45 concrete words that met our frequency threshold of 10 000 uses and were assigned concreteness ratings below 350 (for abstract words) and above 550 (for concrete words). Statistics of the concreteness ratings for the four sets of words are presented in Table 1. The concrete words from both the AB and LG lists have significantly higher concreteness ratings than their corresponding abstract words [$t(87)=-32.8$, $p<.001$; $t(88)=-56.5$, $p<.0001$ for AB and LG, respectively].

In constructing these lists, we cover two bases. With the AB list, we have words that are known to produce concreteness effects as defined in the psycholinguistics literature and allow a fair comparison between our results and those of Audet and Burgess (1999). And with the LG list, we have words that are more grammatically similar and more reliably abstract and concrete. Unintentionally, there is some overlap between the lists. 26 of the 90 LG words are contained in the AB list, 13 abstract and 13 concrete.

Finally, to test for possibly confounding effects of frequency on our measures, we computed the context density and entropy for each of our target words using every single occurrence of the word ("unlimited sampling"), and only the first 5 000 occurrences ("limited sampling").

List	*Word type*	*Mean concreteness rating (std. dev.)*	*Minimum rating*	*Maximum rating*
AB	Abstract	329.9 (40.8)	254	397
AB	Concrete	585.9 (32.4)	425	630
LG	Abstract	305.7 (26.7)	250	350
LG	Concrete	586.3 (19.9)	551	624

Table 1. Concreteness ratings for the two word lists

3. Results

Frequency

A two-way ANOVA finds a main effect of word type on log frequency [$F(1,176)=16.65$, $p=.0001$] and an interaction between word type and list [$F(1,176)=8.29$, $p=.0045$]. That is, when the abstract words from both lists are combined, they are found to be more frequent in our corpus. Recall that Audet and Burgess (1999) also produced this result. However, a post-hoc *t*-test shows that *only* the AB abstract words are significantly more frequent than AB concrete words [$t(88)=4.29$, $p<.0005$]. Thus, the LG list not only controls for frequency in setting a minimum threshold for inclusion, it also balances the frequency between abstract and concrete words.

Context density

To test for effects of word type on context density, two-way ANOVAs were run separately on context density measures computed from unlimited (all occurrences of the word in the corpus) and limited (first 5 000 occurrences) sampling. The mean context densities for each subset of words are presented in Figure 1. Using unlimited sampling, there is a main effect of word type [$F(1,176)=11.29$, $p=.0001$] and an interaction between word type and list [$F(1,176)=5.706$, $p=.02$]. Post hoc *t*-tests found that abstract words have a significantly greater context density than concrete words only for the AB list [$t(88)=3.6203$, $p=.0005$]. Audet and Burgess (1999) achieved this same result.

However, when we limit each word to 5 000 samples, we find the opposite result. There is still a main effect for word type [$F(1,176)=5.35$, $p=.022$], but concrete words now have a *higher* context density than abstract words. The fact that controlling for frequency qualitatively changes the results is not surprising given the very high correlation between context density and log frequency (Figure 2). By contrast, entropy is unaffected by frequency.

Entropy

The mean entropies for the co-occurrence vectors of each subset of words are presented in Figure 3. Again, two-way ANOVAs were performed separately on entropy measures computed from unlimited and limited sampling. Using

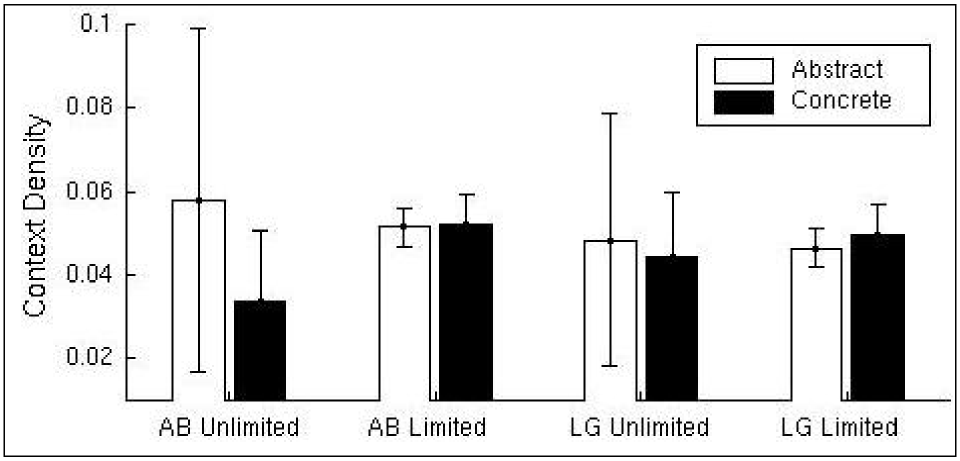

Figure 1. Context density measures for the two word lists, in the limited and unlimited sampling cases.

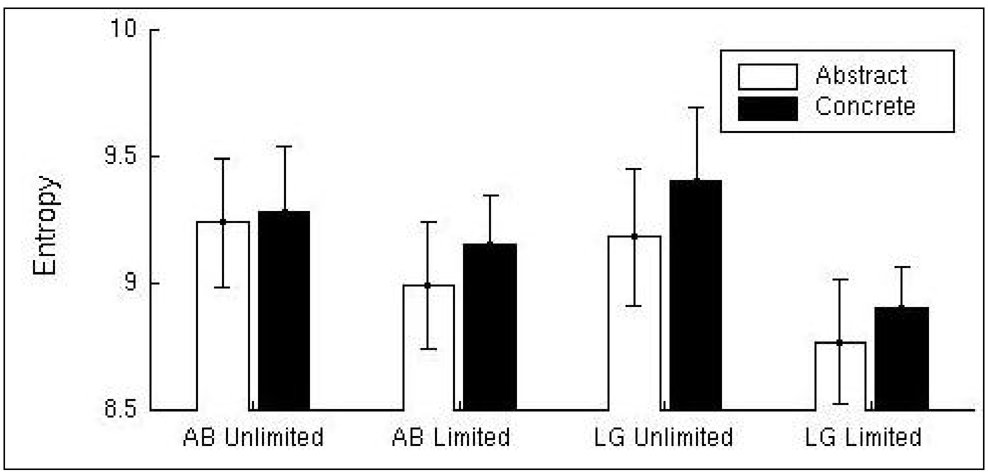

Figure 2. Entropy measures for the two word lists, in the limited and unlimited sampling cases

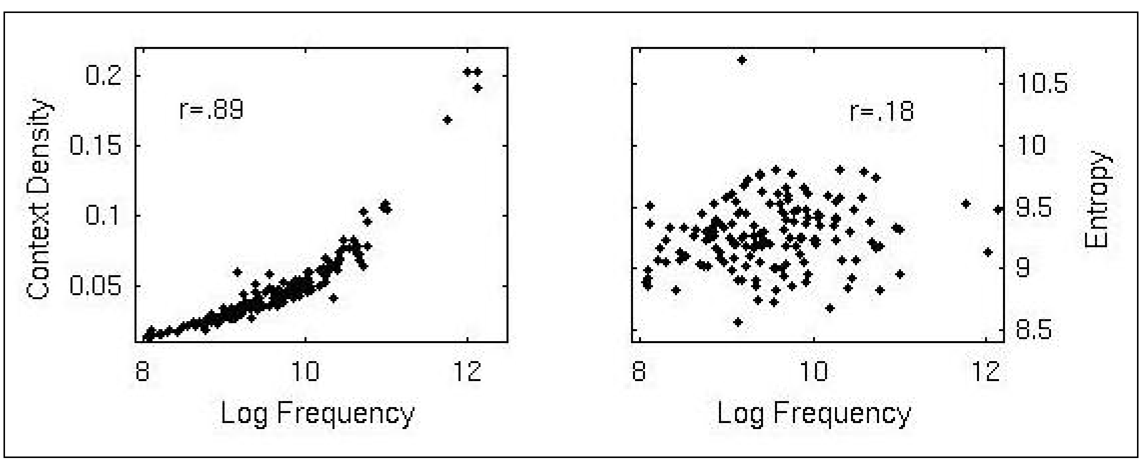

Figure 3. Log frequency correlation plots for context density and entropy.

unlimited sampling, there is a main effect of word type [$F(1,176)=10.39$, $p=.002$] and an interaction between word type and list [$F(1,176)=4.88$, $p=.029$]. Concrete words have a higher entropy than abstract words for both lists, but post hoc *t*-tests find that this difference is significant only for the LG list [$t(88)=-3.683$, $p=.0004$].

Using 5 000 samples of each word does not strongly affect the results. There is still a main effect of type [$F(1,176)=54.07$, $p<.0001$] and now a main effect of word list [$F(1,176)=20.4$, $p<.0001$]. Concrete words produce a greater mean entropy in both lists, and the AB words have a greater mean entropy than LG words. The consistency of entropy measures across the limited and unlimited samplings is to be expected, given its very low correlation with frequency (Figure 2). The unlimited case is therefore a more accurate estimate of entropy.

4. General Discussion

From our findings, we can draw two main conclusions. First, counter to Schwanenflugel's second hypothesis (1991), abstract words appear to be more contextually constrained than concrete words. And second, the work of Audet and Burgess (1999) in support of Schwanenflugel is confounded by effects of word frequency.

The first conclusion stems from the greater entropy of concrete words relative to abstract words in the LG list. However, the fact that the concrete and abstract words in the AB list only show a trend in this direction indicates that this difference, if real, is not terrifically robust. This could be due to part of speech differences. The AB list contains words that are clearly not nouns, while the LG list contains only words that are most commonly used as nouns and no attempt was made to balance the number and type of non-noun words across the abstract and concrete sets. For a discussion on why part of speech must be considered when evaluating a word's context, see Miller and Charles (1991).

Support for the second conclusion is stronger. Audet and Burgess found that abstract words have a higher context density than concrete words. We replicated their findings using a subset of their words but also found that context density is highly correlated with frequency. When we controlled for frequency by uniformly limiting the sampling of each word, we found that concrete words either have a higher context density than abstract words. In light of this, we believe that the Audet and Burgess result is due primarily to the fact that their abstract words were significantly more frequent than their concrete words.

Clearly our results need to be replicated using better controlled word lists with equivalent grammatical properties and multiple, more diverse corpora before we can reliably conclude that concrete words are used in a greater variety of contexts than abstract words. However, these tentative, counterintuitive results are intriguing and, if true, require a rethinking of context availability. How could it be that concrete words, for which contexts are typically easier to generate, are less contextually constrained than abstract words?

One possibility is that although concrete words have weaker, more diffuse associations to contexts, their typical contexts may be more highly associated with one another. Such strong inter-contextual associations could aid in context generation and overcome the handicap of weak word-to-context associations. To illustrate, consider the word "daughter," which should co-occur with contextual elements like "family", "home", "child-rearing", etc. These elements should in turn co-occur with one another, creating a tightly inter-associated set of contexts that might facilitate context generation. Such an effect of third-order associations is not without precedent. McRae et al. (1997) measured the inter-correlations between the features of living things and artifacts and found that it is easier to access the features of a concept that are highly inter-correlated with other features of that concept (e.g. the feature "has fur" is highly correlated with the features "has a tail" and "has four legs").

We have no reason to believe *a priori* that concrete words have stronger inter-contextual associations than abstract words but it is a testable possibility. By computing word co-occurrence or by asking subjects to generate contexts for words, the inter-correlations between contextual elements can be estimated and their relationship to context availability ratings evaluated.

A second possibility is that the greater context availability for concrete words has little to do with how they are used and primarily reflects qualitative differences in their meanings. For example, the *dual-coding* theory of semantic processing (see Paivio, 1991 for a review) posits that people are able to process concrete words with both a linguistic and an *imagistic* semantic system, whereas our processing of abstract words is typically limited to the linguistic system.

In this framework, additional use of the imagistic system facilitates memory for and comprehension of concrete words. Presumably, it could also aid in the generation of contexts for concrete words. As concreteness and context availability ratings are somewhat independent, such a theory cannot completely explain differences in context availability, but the physical features of concrete concepts concrete could be a major factor. This is a possibility that cannot be ruled out until there is evidence from the way concrete and abstract words behave in corpora that explains differences in context availability.

5. Conclusions and future work

In this paper, we tested Schwanenflugel's (1991) hypothesis that abstract words appear in a greater variety of contexts than concrete words by estimating the entropy of the co-occurrence distribution for sets of abstract and concrete words in a corpus of Usenet text. Our evidence suggests that the opposite is true: concrete words tend to be less contextually constrained than abstract words. Furthermore, we showed that previous work (Audet and Burgess 1999) which found support for the hypothesis using the measure of context density to be confounded by frequency and unreliable.

Our counterintuitive results need to be replicated. We plan to repeat our work using words normed for grammatical usage using an automatic part of speech tagger and a larger, more diverse corpus, the British National Corpus as it contains many different registers of spoken and written text. If our results hold, they present an intriguing puzzle for psycholinguistics and demonstrate the importance of testing our intuitions of how words behave using corpus analysis.

References

Akin C 1989 *Context availability and the developmental use of imagery*. Unpublished PhD thesis, Univeristy of Georgia.

Akin C, Schwanenflugel P 1990 *Context availability and developmental trends in the processing of abstract and concrete words*. Unpublished manuscript, University of Georgia.

Altarriba J, Bauer L, Benvenuto C 1999 Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers* 31(4): 578-602.

Audet C, Burgess C. 1999 Using a high-dimensional memory model to evaluate the properties of abstract and concrete words. *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, Vancouver, BC, Canada, pp 37-42.

Bleasdale F 1987 Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Language, Memory, and Cognition* 13: 582-594.

Chiarello C, Senehi J, Nuding S 1987 Semantic priming with abstract and concrete words: Differential asymmetry may be postlexical. *Brain and Language* 31: 43-60.

Cover T, Thomas J 1991 *Elements of information theory*. New York, John Wiley & Sons.

Funnell E 2000 Models of semantic memory. In Best W, Bryan K, and Maxim J (eds) *Semantic Processing: Theory and Practice*. Whurr Publishers, Philadelphia, PA.

Galbraith R, Underwood B 1973 Perceived frequency of concrete and abstract words. *Memory & Cognition* 1:56-60.

Lund K, Burgess C 1996 Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers* 28: 203-208.

Levy J, Bullinaria J 2001 Learning lexical properties from word usage patterns: which context words should be used? In French R, Sougne J (eds), *Connectionist models of learning, development and evolution*. London, Springer, pp 273-282.

McRae K, de Sa V, Seidenberg M 1997 On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General* 126(2): 99-130.

Miller G, Charles W 1991 Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1): 1-28.

Nelson D, McEvoy C, Schreiber T 1998 *The University of South Florida word association, rhyme, and word fragment norms.* http://www.usf.edu/FreeAssocation.

Paivio A 1986 Mental representations: A dual coding approach. Oxford, England: Oxford University Press.

Paivio A 1991 Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45: 255-287.

Schwanenflugel P 1991 Why are abstract concepts hard to understand? In Schwanenflugel P (ed), *The psychology of word meanings*. Hillsdale, NJ, Lawrence Erlbaum Associates, pp 223-250.

Schwanenflugel P, Harnishfeger K, Stowe R 1988 Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language* 27: 499-520.

Schwanenflugel P, Shoben, E 1983 Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory & Cognition* 9: 82-102.

Schwanenflugel P, Stowe R 1989 Context availability and the processing of abstract and concrete words. *Reading Research Quarterly* 24: 114-126.

Wiemer-Hastings K, Graesser A 2000 Contextually representing abstract concepts with abstract structures. *Poster presented at the 22nd Annual Conference of the Cognitive Science Society*. Philadelphia, PA.

Wilson M 1988 The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers* 20: 6-11.