Title
Testing the Distributioanl Hypothesis: The influence of Context on Judgements of Semantic Similarity

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)

ISSN
1069-7977

Authors
McDonald, Scott
Ramscar, Michael

Publication Date
2001

Peer reviewed

Testing the Distributional Hypothesis: The Influence of Context on Judgements of Semantic Similarity

Scott McDonald (scottm@cogsci.ed.ac.uk)
Michael Ramscar (michael@cogsci.ed.ac.uk)
Institute for Communicating and Collaborative Systems, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW Scotland

Abstract

Distributional information has recently been implicated as playing an important role in several aspects of language ability. Learning the meaning of a word is thought to be dependent, at least in part, on exposure to the word in its linguistic contexts of use. In two experiments, we manipulated subjects' contextual experience with marginally familiar and nonce words. Results showed that similarity judgements involving these words were affected by the distributional properties of the contexts in which they were read. The accrual of contextual experience was simulated in a semantic space model, by successively adding larger amounts of experience in the form of item-in-context exemplars sampled from the British National Corpus. The experiments and the simulation provide support for the role of distributional information in developing representations of word meaning.

The Distributional Hypothesis

The basic human ability of language understanding – making sense of another person's utterances – does not develop in isolation from the environment. There is a growing body of research suggesting that *distributional information* plays a more powerful role than previously thought in a number of aspects of language processing. The exploitation of statistical regularities in the linguistic environment has been put forward to explain how language learners accomplish tasks from segmenting speech to bootstrapping word meaning. For example, Saffran, Aslin and Newport (1996) have demonstrated that infants are highly sensitive to simple conditional probability statistics, indicating how the ability to segment the speech stream into words may be realised. Adults, when faced with the task of identifying the word boundaries in an artificial language, also appear able to readily exploit such statistics (Saffran, Newport & Aslin, 1996). Redington, Chater and Finch (1998) have proposed that distributional information may contribute to the acquisition of syntactic knowledge by children. Useful information about the similarities and differences in the meaning of words has also been shown to be present in simple distributional statistics (e.g., Landauer & Dumais, 1997; McDonald, 2000).

Based on the convergence of these recent studies into a cognitive role for distributional information in explaining language ability, we call the general principle under exploration the *Distributional Hypothesis*. The purpose of the present paper is to further test the distributional hypothesis, by examining the influence of context on similarity judgements involving marginally familiar and novel words. Our investigations are framed under the 'semantic space' approach to representing word meaning, to which we turn next.

Distributional Models of Word Meaning

The distributional hypothesis has provided the motivation for a class of objective statistical methods for representing meaning. Although the surge of interest in the approach arose in the fields of computational linguistics and information retrieval (e.g., Schutze, 1998; Grefenstette, 1994), where large-scale models of lexical semantics are crucial for tasks such as word sense disambiguation, high-dimensional 'semantic space' models are also useful tools for investigating how the brain represents the meaning of words.

Word meaning can be considered to vary along many dimensions; semantic space models attempt to capture this variation in a coherent way, by positioning words in a geometric space. How to determine what the crucial dimensions are has been a long-standing problem; a recent and fruitful approach to this issue has been to label the dimensions of semantic space with *words*. A word is located in the space according to the degree to which it *co-occurs* with each of the words labelling the dimensions of the space. Co-occurrence frequency information is extracted from a record of language experience – a large corpus of natural language. Using this approach, two words that tend to occur in similar linguistic contexts – that is, they are *distributionally* similar – will be positioned closer together in semantic space than two words which are not as distributionally similar. Such simple distributional knowledge has been implicated in a variety of language processing behaviours, such as lexical priming (e.g., Lowe & McDonald, 2000; Lund, Burgess & Atchley, 1995; McDonald & Lowe, 1998), synonym selection (Landauer & Dumais, 1997), retrieval in analogical reasoning (Ramscar & Yarlett, 2000) and judgements of semantic similarity (McDonald, 2000).

Contextual co-occurrence, the fundamental relationship underlying the success of the semantic space approach to representing word meaning, can be defined in a number of ways. Perhaps the simplest (and the approach taken in the majority of the studies cited above) is to define co-occurrence in terms of a 'context window': the co-occurrence frequency of w_1 with w_2 is defined as the number of times that w_2 (the 'context word') occurs in the window of n words surrounding w_1, summed over all instances of w_1 in the corpus. Given a set of k context words, any word in the vocabulary can be represented as a k-dimensional vector of co-occurrence frequencies. The best fit to psychological data is typically achieved with word vectors constructed using context window sizes between ±2 and ±10 words (see, e.g., Patel, Bullinaria & Levy, 1998).

Besides its emphasis on identifying a potential source of information useful for the development of semantic representations, the distributional hypothesis also accommodates predictions about the consequences of *manipulating* the learning environment. By modifying the degree of distributional similarity holding between two words in a person's language experience, a particular word's location in semantic space can be adjusted (i.e., a word vector can be 'pushed' in a given direction). In Experiments 1 and 2 we test whether manipulating contextual co-occurrence has behavioural consequences, by eliciting judgements of semantic similarity involving marginally familiar and nonce words embedded in biasing contexts.

Learning Word Meaning from Context

It is well-established that the context in which an unfamiliar word occurs is an important determinant of how much is learned about the word, and it is apparent that context often provides the sole means for establishing its meaning (e.g., Carnine, Kameenui & Coyle, 1994; Fischer, 1994). In order to interpret an unknown word, the context provides cues, in the form of some combination of: (1) the identity of the words in the context surrounding the unknown word and the relationships between these words and the unknown word (i.e., distributional information); (2) world knowledge retrieved from long-term memory associated with these words; and (3) the cognitive model of the discourse (or situation) currently being built. But it seems that distributional information on its own, if suitably constraining, could be sufficient for determining the meaning of an unfamiliar word. Consider the occurrence of the neologism *broamed* in the following context:

Because the capsule was hermetically broamed, its contents were in perfect condition after more than a hundred years under water.

In this example, knowledge about the distributional behaviour of *hermetically* certainly guides the inference that the meaning of *broamed* is similar to the meaning of *sealed*, because *hermetically* nearly always co-occurs with *sealed*. Further support for this inference is contributed by knowledge about capsules and the conditions required in order for something to remain in perfect condition in adverse circumstances.

Contextual cues also play an important role in consolidating the meaning of newly-learned words. The more exemplars of a word in its context of use that are encountered, the more its meaning can be refined and delimited, especially if one has some prior knowledge of the discourse or passage topic. We assume that a close correspondence exists between a word's subjective familiarity and the amount of experience one has with the word. The less experience, the less familiar the word and the less established its semantic representation in the brain.

In the experiments reported below, we attempt to manipulate the distributional knowledge associated with sets of marginally familiar and completely novel words in order to test a basic prediction of semantic space models in particular and the distributional hypothesis in general. Distributional information is the only variable manipulated; for each item we constructed two different paragraph contexts, each containing only four exemplars of the item. By judicious selection of the words in the context surrounding each instance of the word of interest, co-occurrence patterns can be created that resemble the patterns of other, more familiar words. Using semantic space model terminology, a word vector can be 'pushed' towards another vector by bringing dimensions of the space into alignment. The question we addressed was whether this manipulation of distributional information was sufficient to influence subjects' ratings of semantic similarity.

Experiment 1

Experiment 1 focuses on marginally familiar words. These are words that one is likely to have encountered, but not with sufficient frequency to have a firm grasp of their meaning. For instance, one might know that a *samovar* is some kind of utensil associated with hot drinks, but be unsure about whether it is used for making the drink or for serving it. So one might be equally willing to accept that *samovar* signifies something like a kettle or an urn. By exposing subjects to paragraphs containing exemplars of *samovar* together with contextual cues lexically associated with each of these possible interpretations (i.e., *urn* vs. *kettle*), subjects' representations of the meaning of *samovar* may be nudged towards the meaning of the word associated with the contextual cues. Thus the dependent variable we would like to measure is the similarity of the two words' semantic representations.

While such a measurement is not directly possible, psychologists have developed a number of indirect methods that purport to tap into the semantic representations of words. We needed a task that would allow similarity in meaning to be reliably measured, while at the same time remain sensitive to the hypothesised changes in semantic representations due to the context manipulation. Similarity ratings meet these criteria, having a long history of use in psychological investigations of word meaning (e.g., Osgoode, Suci & Tannenbaum, 1957), and importantly, similarity judgements have been shown to be affected by context. For instance, Barsalou (1982) demonstrated that in a 'pets' context, the concepts *snake* and *raccoon* were judged to be more similar than if no context was provided. Medin, Goldstone and Gentner (1993) also observed context-dependent similarity effects: *black* was rated as more similar to *white* when also compared to *red* than when *black*⇔*white* was the only comparison required. We expected that subjects' ratings of between-word similarity, such as *samovar*⇔*kettle*, would be

Context A: 'urn'

On his recent holiday in Ghazistan, Joe slipped easily into the customs of the locals. In the hotel restaurant there was a samovar dispensing tea at every table. Guests simply served themselves from the samovar whenever they liked. Joe's table had an elaborately crafted samovar. It was the first earthenware samovar that he had seen.

Context B: 'kettle'

On his recent holiday in Ghazistan, Joe slipped easily into the customs of the locals. His hotel room featured a samovar and a single hob. Each morning Joe boiled water in the samovar for tea. Like others he had seen on his holiday, Joe's samovar was blackened from years of use. He imagined that at some point it would be replaced with an electric samovar.

Figure 1. The *urn*-biased and *kettle*-biased paragraph contexts created for *samovar*.

similarly influenced by the properties of the paragraph context which they had just read.

Method

Participants Forty-eight subjects, mostly undergraduate Psychology students at the University of Edinburgh, were recruited. All participants were native speakers of British English.

Materials and Design A list of 20 marginally familiar words (ten nouns and ten verbs) was compiled. Sixteen items were selected from the pre-tested materials used by Chaffin (1997) in his study examining free associations made to high- and low-familiarity words, and the remaining four were chosen by the authors. Items ranged in frequency from 0.13 to 2.92 occurrences per million (median: 0.64), according to a lemma frequency list created from the 100 million word British National Corpus (BNC).

For each item, we generated two 'target meanings' which we felt were plausible interpretations of the items. Then, for each of these target meanings we composed a short paragraph containing exactly four exemplars of the item. (See Figure 1 for a representative item with its paragraph contexts). Text passages were homogenous in structure, with the first sentence setting the scene; the marginally familiar words were embedded in the following three or four sentences. Passages ranged in length from 50 to 96 words (median length of 62). We attempted to bias the interpretation of the item in the paragraph by seeding the immediate context of each exemplar with strong lexical associates of the selected target meaning. For example, the meaning of *samovar* in Context B is 'pushed' towards *kettle* through the words *boiled*, *blackened* and *electric*, which are all more indicative of kettles than urns.

The strong lexical associates were generated in turn using a statistical technique commonly employed in computational linguistics for discovering collocations (e.g., Church & Hanks, 1990; Manning & Schütze, 1999); this procedure involved, for each target meaning (e.g., *urn*, *kettle*), collecting the co-occurrence frequencies of all words found in a ±5 word window around it in the BNC, converting these counts using the log-transformed odds ratio statistic (Agresti, 1990), and then sorting the resulting list. Strong associates – roughly, words that co-occur more often than expected by chance – tend to appear at the top of the ranking. We then selected suitable words for use as contextual cues from the topmost part of the list.

Paragraph contexts were randomly assigned to one of the two levels of the Context factor (A, B). This design is now sufficient to test for an effect of Context when subjects are asked to rate the similarity between e.g., *samovar* and *urn* after reading either Context A or Context B. In order to complete a factorial design, Context was crossed with a second factor, Target Meaning, with the same two levels, varying the word to which the marginally familiar item is compared.

The materials were next divided into four versions of 20 paragraphs each. Counterbalancing ensured that no participant saw the same item more than once.

Procedure Subjects were divided randomly amongst each of the four versions. The experiment was administered in the form of a questionnaire, with one paragraph context per page. Located below each paragraph was a numbered seven-point scale, and subjects were instructed to rate how similar the item was to the target meaning, where 'a 1 means "not at all similar" and a 7 means "highly similar"'; e.g., "How similar is a *samovar* to an *urn*?". The verb items were presented in present participle form; e.g., "How similar is *absconding* to *escaping?*". Order of presentation of the 20 items was randomised individually for each participant.

After completing the 20 items, subjects were required to rate a list of 28 words for familiarity, also using a 7-point scale, where 'a 1 means "very unfamiliar" and a 7 means "very familiar"'. This list comprised the 20 designated items plus eight filler words of moderate to high familiarity. The purpose of the familiarity ratings task was to allow a more detailed examination of the similarity data, in order to take into consideration the inherent variability in individuals' experience with the items.

Results

We conducted two-way repeated measures analyses of variance (ANOVAs) on the similarity judgements, treating both subjects and items as random factors.

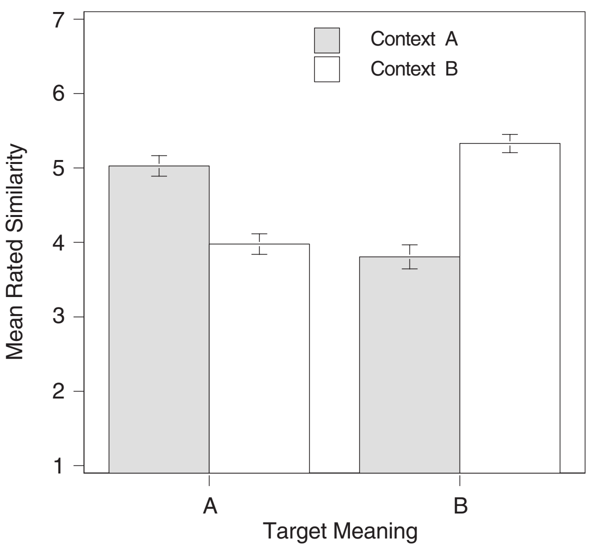

Figure 2. Mean semantic similarity as a function of Context and Target Meaning in Experiment 1.

There were no reliable main effects of either Target Meaning, $F_1(1,47)=4.02$, $MSE=0.667$, $p>0.05$; $F_2(1,19)=1.59$, $MSE=0.694$, $p>0.2$, or Context, $F_1(1,47)<1$; $F_2(1,19)<1$. The lack of a Target Meaning main effect indicates that, collapsing over the paragraph contexts in which the marginally familiar items were embedded, there was no bias between the 'A' and 'B' meanings in terms of their rated similarity to the item. The lack of a main effect of Context indicates an analogous absence of bias for the paragraph contexts.

There was a highly significant Context × Target Meaning interaction: $F_1(1,47)=60.04$, $MSE=1.323$, $p<0.001$; $F_2(1,19)=35.73$, $MSE=0.924$, $p<0.001$). As indicated by Figure 2, the interaction was due to Context effects at each level of Target Meaning. The mean similarity rating between a marginally familiar word and its 'A' meaning was higher when the item was embedded in the context biasing that meaning than when it appeared in the passage biasing the 'B' meaning.

Discussion

These results indicate that the distributional information contained in the paragraph contexts are sufficient to influence participants' similarity judgements. In the terminology of semantic space models, vectors were successfully 'pushed' towards other vectors in the representational space. Thus a strong prediction of the semantic space theory of meaning representation is supported: by selecting appropriate contextual cues and positioning them in the immediate linguistic context of a marginally familiar word, behavioural measures assumed to tap the word's meaningful properties can be influenced.

The results also provide support for the distributional hypothesis. Adding instances of a word in its environment of use to one's language experience – even as few as four exemplars – appears to be adequate to affect one's perception of its similarity in meaning to other words.

Although the items were chosen to be on the frontiers of familiarity for the subject population, the familiarity of a particular word can vary substantially between participants. For example, *samovar* may be a familiar word to someone who has travelled in Russia. According to the distributional hypothesis, this individual should be less influenced by the context when rating the similarity of *samovar* to *kettle* or to *urn*.

As we had collected familiarity ratings for each of the targets from each subject, we were able to address this question by dividing the ratings data points into low-familiarity (LoFam) and high-familiarity (HiFam) groups around the median familiarity score. The LoFam partition included data points with a self-rated familiarity score of three or less, and the HiFam group contained data for items rated as five or more.

The critical Context × Target Meaning interaction was present in the LoFam partition: $F_1(1,29)=59.24$, $MSE=1.80$, $p<0.001$; $F_2(1,17)=21.61$, $MSE=1.82$, $p<0.001$. The HiFam partition also displayed the interaction: $F_1(1,36)=21.55$, $MSE=1.80$, $p<0.001$; $F_2(1,17)= 30.28$, $MSE=0.92$, $p<0.001$.

It seems, then, that subjects' interpretations of marginally familiar words could be guided by the distributional properties of the contexts in which they were encountered, at least to the extent necessary to influence an immediately executed similarity rating. This effect was observed both for words with which subjects considered themselves reasonably familiar and for less familiar words.

The results of Experiment 1 raise two interesting questions with regard to our subjects' mental representations of the meanings of the stimuli: Were subjects actively using the distributional information in the contexts to actively augment (or even construct) their representation of the meaning of *samovar*? Or were the paragraph contexts activating particular features of their existing knowledge about samovars, causing the attendant shift in similarity ratings? In the latter case it could be argued that subjects' sensitivity to the distributional properties of words demonstrated in Experiment 1 is merely an epiphenomenon, a reflection of the fact that certain concepts share certain semantic features. On this account, the distributional properties associated with words arise *because* the concepts underlying the words possess certain features, and it is sensitivity to similarities between these concepts that subjects are actually manifesting. To examine these competing explanations, Experiment 2 controlled for the influence of any such prior conceptual knowledge by replacing Experiment 1's items with nonce words. Subjects were essentially starting from a 'tabula rasa' with respect to the meaning of nonce words, so evidence that the context was truly exerting an independent influence on subjects' judgements in Experiment 1 would be provided if similar effects of context are observed using nonce words.

Experiment 2

Experiment 2 controlled for the potential influence of participants' existing conceptual knowledge about the meaning of the target items by replacing the marginally familiar items used in Experiment 1 with nonce words. (Thus the task now closely resembles the situation where an unknown word is encountered during reading, and its meaning has to be inferred from the context.)

Method

Participants Twenty subjects from the same population as Experiment 1 volunteered to take part.

Materials and Design The materials were identical to those used in Experiment 1, with the exception that the 20 marginally familiar items were replaced with orthographically-legal and pronounceable nonwords. For instance, all occurrences of *samovar* in the text passages were replaced with the nonce word *balak*. Care was taken that each nonce replacement did not phonologically resemble the original item or its two associated 'target meanings'.

Procedure The procedure was the same as for Experiment 1, except there was no familiarity ratings task.

Results and Discussion

Similarity ratings data were submitted to repeated measures ANOVAs. The Target Meaning × Context interaction was significant both by subjects: $F_1(1,19)=159.83$, $MSE=0.469$, $p<0.001$; and by items: $F_2(1,19)=40.23$, $MSE=1.863$, $p<0.001$. There were no main effects of either Target Meaning: $F_1(1,19)=1.09$, $MSE=0.385$, $p>0.3$; $F_2(1,19)<1$ or Context: $F_1(1,19)<1$; $F_2(1,19)<1$.

Thus these results are consistent with the findings of Experiment 1. It appears that any objections regarding the possible role and influence of prior knowledge about the meanings of Experiment 1's marginally familiar items are unfounded. Similarity comparisons involving unknown (nonce) words were also susceptible to manipulation of the same contextual cues that gave rise to the interaction in Experiment 1.

Simulating the Accumulation of Contextual Experience

Experiments 1 and 2 have shown that a very small amount of experience with a word in context is capable of influencing similarity judgements involving that word. The items in Experiment 1 were selected to represent the sorts of words to which subjects would be expected to have a low level of prior exposure. If it were possible to *increase* the amount of one's prior contextual experience with a given item, the influence of subsequent exposure (i.e., the four-exemplar paragraphs in Experiment 1) should be reduced. We simulated this effect of previous experience using a semantic space model derived from distributional statistics. We predicted that the size of the simulated context effect would diminish as the ratio of previous experience to the experience provided by the paragraphs increased. We varied the amount of contextual exposure given to the model by varying the size of the corpus used to construct co-occurrence vector representations for the 20 marginally familiar items.

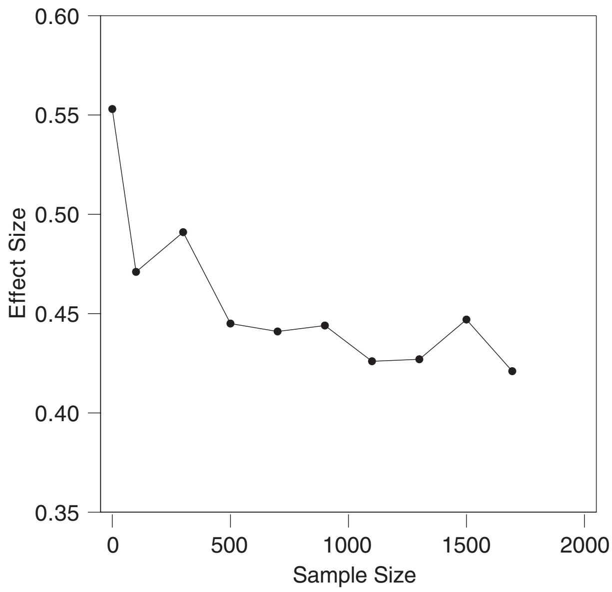

Figure 3. The size of the Consistency effect as a function of the amount of contextual experience.

Method

From the BNC, we extracted the ±5 word contexts surrounding every occurrence of all 20 items (a total of 1,694). We then took random samples (with replacement) of various sizes from this item-in-context 'corpus', appending them to both an analogous corpus formed by the 'A' passages and the corpus formed by the 'B' paragraphs, resulting in separate 'A' and 'B' corpora for each sample size.

From each 'corpus', we extracted co-occurrence vectors for the 20 items using a window size of ±5 words and the 20,000 most frequent content words as context words. The resulting item vectors thus directly reflect the ratio of previous experience to subsequent experience (vectors created from the passages only simulate a complete lack of previous experience with the word). Vectors for the 40 'target meanings' (e.g., *urn*, *kettle*) were constructed using the entire BNC.

Results and Discussion

We collapsed the 2 × 2 design of Experiment 1 into a single factor, Consistency, in order to compare the vector similarity of an item with each of its 'target meanings', between the case where the paragraph context is consistent with (or biases) the target meaning (e.g., *samovar⇔urn* for Context 'A'; see Figure 1) and the case where it is inconsistent (*samovar⇔urn* for Context 'B'). Similarity was computed as the cosine of the angle between vectors, and a paired-*t* test was conducted on the cosine measurements. Consistent comparisons should return a larger cosine than Inconsistent comparisons. At the $\alpha=0.01$ level of significance, reliable Consistency effects were observed for all sample sizes but one (the effect for the 1100-exemplar sample was significant at $\alpha=0.05$).

In order to illustrate the effect of increasing the amount of previous experience, Figure 3 displays the Consistency effect size (Cohen's *d*) as the sample size varies. As expected, the effect is largest for vectors created from the passages only, and diminishes as more contextual experience is added. Both Experiment 1's results and the anticipated effect of variable amounts of prior exposure were simulated in a semantic space model drawing only upon distributional information.

General Discussion

To summarise, manipulating the contextual cues present in short text passages was sufficient to influence adults' similarity judgements involving marginally familiar and nonce words embedded in these passages. Our results suggest that readers' interpretations of these items were 'pushed' towards the meanings of other words. Analogous to the way that the meaning of unknown words can be determined while reading, contextual information is also an influential factor when consolidating the meaning of words on the frontiers of familiarity.

The experimental results also suggest that a remarkably small amount of exposure to a word in a meaningful context is sufficient to influence similarity ratings. However, the relative recency of this experience is likely an important factor; the context effect may well diminish as a function of the length of time between reading the paragraph and making the similarity judgement.

Though a simple model of word learning, the semantic space simulation illustrated the decrease in susceptibility to contextual manipulation expected as one's prior experience with a word increases. Of course, we do not claim that human semantic space has 20,000 dimensions; rather, what is important is the inferences that can be drawn about a word's meaning simply by taking note of the words in its immediate context. It is notable that the simulated Consistency effect was still reliable even after all the contextual experience in the BNC was added; in as much as the BNC can be considered to represent the average person's language exposure, it seems that very little extra contextual experience is needed to affect the perception of a word's similarity in meaning to other words.

Acknowledgements

We thank Dan Yarlett for useful discussion and comments.

References

Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82-93.

Carnine, D., Kameenui, E. J. & Coyle, G. (1984). Utilisation of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, *19*, 188-204.

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition*, *25*, 203-226.

Church, K. W. & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*, 22-29.

Fischer, U. (1994). Learning words from context and dictionaries: an experimental approach. *Applied Psycholinguistics*, *15*, 551-574.

Grefenstette, G. (1994). *Explorations in automatic thesaurus discovery*. Boston: Kluwer.

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.

Lowe, W. & McDonald, S. (2000). The direct route: Mediated priming in semantic space. *Proceedings of the 22th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Lund, K., Burgess, C. & Atchley, R. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 660-665). Mahwah, NJ: Erlbaum.

Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press

McDonald, S. (2000). *Environmental determinants of lexical processing effort*. Doctoral dissertation, Division of Informatics, University of Edinburgh.

McDonald, S. & Lowe, W. (1998). Modelling functional priming and the associative boost. *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 667-680). Mahwah, NJ: Erlbaum.

Medin, D. L., Goldstone, R. L. & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254-278.

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

Patel, M., Bullinaria, J. A. & Levy, J. P. (1998). Extracting semantic representations from large text corpora. In J. A. Bullinaria, D. Glasspool & G. Houghton (Eds.) *Proceedings of the 4th Neural Computation and Psychology Workshop, London, 9-11 April 1997*. London: Springer-Verlag.

Ramscar, M. J. A. & Yarlett, D. G. (2000). The use of a high-dimensional, "environmental" context space to model retrieval in analogy and similarity-based transfer. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 381-386). Mahwah, NJ: Erlbaum

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425-469.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Saffran, J. R., Newport, E. L. & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606-621.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*, 97-124.