Resolution of Inferential Descriptions in Lexical Clusters

Sabine Schulte im Walde
University of Stuttgart
IMS
Azenbergstr. 12
70174 Stuttgart
Germany
schulte@ims.uni-stuttgart.de

Massimo Poesio
University of Edinburgh
Cognitive Science / HCRC
2 Buccleuch Place
Edinburgh EH8 9LW
Scotland, UK
poesio@cogsci.ed.ac.uk

Chris Brew
University of Edinburgh
LTG
2 Buccleuch Place
Edinburgh EH8 9LW
Scotland, UK
chrisbr@cogsci.ed.ac.uk

Abstract

We present preliminary results concerning the use of lexical clustering algorithms to acquire the kind of lexical knowledge needed to resolve definite descriptions, and in particular what we call 'inferential' descriptions. We tested the hypothesis that the antecedent of an inferential description is primarily identified on the basis of its semantic distance from the description. We used various parameters for the co-occurrence clustering algorithm and different approaches to measure the distance between the lexical vectors. We found that in those cases in which the sort of lexical knowledge we acquired is the main factor, the algorithms we used performed reasonably well; however, factors other than semantic distance play the main role in the majority of cases; several standing problems are discussed.

Introduction

In order to develop systems for anaphoric resolution whose generality and performance can be evaluated in a quantitative fashion–i.e., by testing them over a corpus including texts from different domains–it is necessary to address the issue of commonsense knowledge. The question we are currently studying is what kind of commonsense knowledge is involved in the resolution of so-called BRIDGING DESCRIPTIONS (Clark 1977), i.e., definite descriptions that refer to an object introduced into the common ground as the result of the mention of a related object–such as *the door* in

(1) *John walked towards the house. THE DOOR was open.*

Arguably, the minimal hypothesis to pursue in this connection is that resolving these descriptions is purely a matter of lexical knowledge–i.e., that the identification of the antecedent depends solely on the degree of association among lexical items. The assumption that the lexicon is organised like a 'semantic' network where some concepts are more closely related than others, originally motivated by semantic priming effects (Meyer & Schvaneveldt 1971), underlies most current psychological models of the lexicon, including WordNet (Miller *et al.* 1990) and has been adopted in much research on reference resolution: such models assume that the antecedent for *the door* in (1) is found by looking for an antecedent whose concept is semantically close (in some sense), and that *the house* is chosen because the concept associated with this antecedent subsumes the concept associated with the inferential description in the semantic network. We will call this the MAIN HYPOTHESIS: Resolving an inferential description is a matter of finding the antecedent in the text that primes the head predicate of the inferential description most strongly.

One possibility to get the information needed to test this hypothesis is to use an existing source of lexical knowledge, such as WordNet; however, the results we obtained with this method–reported in (Poesio, Vieira, & Teufel 1997)–were not too satisfactory, owing to the incompleteness of the information hand-coded in WordNet, as well as to several inconsistencies we found in it. As a result, we have been exploring techniques for acquiring this information automatically. In the initial phase, we have mainly been experimenting with clustering algorithms (Charniak 1993).

In particular, the work discussed here was inspired by (Lund, Burgess, & Atchley 1995), who reported that the clusters of words obtained with their HAL model of lexical clustering reflect a notion of distance that correlates well with subjects' results on semantic priming tasks. This work offers therefore the opportunity to test the hypothesis discussed above that resolving inferential descriptions is a matter of semantic priming. We assessed the performance of several variants of the HAL method on this task.

Background

Inferential Descriptions

Our studies of definite description use (Poesio & Vieira 1998; Vieira & Teufel 1997; Poesio, Vieira, & Teufel 1997) led to the development of a taxonomy of definite descriptions reflecting the types of commonsense knowledge that appear to be involved in their resolution. For the purposes of this paper, we will consider definite descriptions as falling in one of the following three categories:[1]

Anaphoric same head: these are the definite descriptions whose resolution involves simply match-

[1] As discussed in (Poesio & Vieira 1998), these categories are not completely mutually exclusive.

ing the head of the antecedent with the head of the definite description, as in *a car ... the car*;[2]

Inferential: this is a semantically eclectic class, including those definite descriptions whose head is not identical to that of the antecedent, and those whose relation with the antecedent is not one of co-reference. This class also includes references to events, as in *John killed Bill. THE MURDER took place at 5pm*, and to entities introduced by proper names, as in *We are celebrating this year 200 years since Franz Schubert's birth. THE FAMOUS COMPOSER was born in 1797.*

Discourse new: this class consists of those definite descriptions that do not have an antecedent in the text, and includes both references to 'larger situation' knowledge such as *the sun* and possible first-mention definite descriptions such as *the first man to sail to America* (Hawkins 1978; Prince 1981; Poesio & Vieira 1998).

In this paper we are exclusively concerned with the class of inferential descriptions, also called 'bridging descriptions' following Clark's taxonomy. In order to categorise the types of commonsense knowledge involved in their resolutions and to gain a feeling for how many of the required inferences would be supported by a semantic network, this class was further analysed in (Vieira & Teufel 1997; Poesio, Vieira, & Teufel 1997); the following categories of inferential descriptions were identified:

Synonymy: the antecedent and the inferential description are synonymous, as in *a new album – the record.*

Hypernymy/Hyponymy: the antecedent and the inferential description are in a *is-a*-relation, as in *rice – the plant* (super-ordination/hypernymy) or *a plant – the rice* (sub-ordination/hyponymy).

Meronymy: the antecedent and the inferential description stand in a *part-of* relation, as in *a tree – the leaves.*

Names: the inferential description refers back to a proper name, as in *Bach – the composer.*

Compound Nouns: the 'antecedent' occurs as part of a compound noun, as in *the stock market crash – the markets.*

Events: the antecedent is not introduced by a noun phrase, but by either a verb phrase or a sentence, e.g. *they planned – the strategy.*

[2]In fact, even resolving these cases may involve some form of commonsense inference–e.g., to take into account the effects of pre-modifiers and post-modifiers in recognising that *a blue car* cannot serve as the antecedent of *the red car* in *I saw a blue car and a red car. The red car was a Ferrari.*

Discourse Topic: the antecedent is the –often implicit–'discourse topic' of a text, as in *the industry* appearing in a text about oil companies.

(General) Inference: the inferential description is based on more complex inferential relations such as causal inferences, as in *last week's earthquake – the suffering people.*

The first three classes include the inferential descriptions whose resolution we might expect to be supported by the sort of information stored in a typical semantic network such as WordNet; these networks also include information about individuals of the kind needed to resolve inferential descriptions in the 'Names' class.

Poesio, Vieira and Teufel ran a test on a corpus of 20 parsed Wall Street Journal articles from the Penn Treebank, including 1040 definite descriptions, of which 204 were classified as inferential. When trying to resolve an inferential description, the discourse entities in the previous five sentences were considered as potential antecedents, and WordNet was queried to find a relation between the inferential description and each antecedent. WordNet found a relation between an inferential description and an antecedent for 107 of these descriptions, but in only 34 cases was the right antecedent suggested. Separate heuristic-based techniques were also proposed, so that in total 77 descriptions were identified correctly.

Acquiring Semantic Networks by Clustering

'Clustering' is a popular approach to lexical acquisition based on the idea that semantically related words are close to each other in some higher-dimensional space representation where they form 'clusters' of similar words–i.e., the very same intuition behind research on semantic networks. Clustering algorithms view each word as a point in an n-dimensional space, i.e., as a vector of size n, and the similarity between words is measured in terms of the distance between the points that represent them. The goal of clustering algorithms is to construct such a representation automatically, exploiting a corpus. These methods differ depending on the dimensions used and their number, on the metric used to measure the distance among the points, and the algorithm used to construct the vectors (Charniak 1993).

A common approach to clustering is to just use words w_i as dimensions–often called CONTEXT WORDS–, i.e., to let the vector associated with word w, $\vec{w}$, be a record of how frequently w occurred close to the words w_i; the intuition is that a word is defined by the 'company that it keeps', i.e., by the words with which it is most frequently encountered. Algorithms which assign words vector representations of this type scan a text and whenever they encounter a word w they increment all cells of $\vec{w}$ corresponding to the words w_i that occur in the vicinity of w. The degree of vicinity is typically defined by a window of fixed size, by the number of either words or characters, varying strongly (compare, for example, a window of 5 words as in (Church & Hanks 1989) with a window of 100 characters as in (Schütze 1992)).

Once the vectors associated with each word have been constructed in this way, we can estimate the semantic similarity between words by measuring the distance between the associated vectors. A great number of distance measures have been suggested, but the following three are the best known:

- **Manhattan Metric:**
  The Manhattan Metric measures the distance of two points in n-dimensional space by summing the absolute differences of the vectors' elements:
  $$d = \sum_{i=1}^{n} |x_i - y_i|$$
- **Euclidean Distance:**
  The Euclidean Distance is calculated by summing the squared differences of the vectors' elements and then determining the square root:
  $$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$
- **Cosine of the Vectors' Angle:**
  This measure does not calculate the distance between points, but the angle α between the n-dimensional vectors which determine the points in n-dimensional space:
  $$cos(\alpha) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}.$$
  The closer the $cos(\alpha)$ is to 1, the smaller the angle α is and therefore the shorter the distance is.

Other measures proposed in the literature include Spearman Rank correlation coefficient, Hellinger distance, and Kullback-Leibler divergence. Weighted combinations of different measures have also been used. (See (Levy, Bullinaria, & Patel 1997) for some discussion.)

Lund et al's HAL Model

Lund et al. (1995) used a 160 million word corpus of articles extracted from all newsgroups containing English dialogue. They chose as context words the 70,000 most frequently occurring symbols within the corpus.

The co-occurrence counts were calculated as follows. They defined a window size of 10 words to the left and to the right of the target words, and within this window, the co-occurrence values were inversely proportional to the number of words separating a specific pair. So, whenever a target word w was encountered, the context vector $\vec{w}$ was incremented as follows: the count $\vec{w}[w_1]$ for the word w_1 next to the target word was incremented by 10, the count $\vec{w}[w_2]$ for the next word was incremented by 9, and so forth, thus weighting the closeness of the co-occurring words.

To reduce the amount of data, the column variances of the particular vectors used in each experiment were computed, and the columns with the smallest variances were discarded. This left a 200-element vector for each target word.

Our Methods

In our experiments we adopted the fundamental aspects from the clustering technique of Lund et al, parameterising several of its aspects in order to evaluate not only the Main Hypothesis, but also the influence of certain parameters on the results. We briefly discuss our methods here; for more discussion and details, see (Schulte im Walde 1997).

As in the case of Lund et al, our basic clustering algorithm involves associating with each word a vector whose dimensions are other words; and again as in their case, the vectors are constructed by scanning a text, considering for each word w that is encountered all neighbours w_i in a window of size n, and increasing by a factor possibly weighted by distance the cells of w's vectors associated with each w_i. This algorithm was made parametric on window size (we considered sizes 1,2,3,5,10,20 and 30), on whether inflected words or their lemmas were considered, and on whether just words or word/tag pairs were used.

We ran some preliminary experiments to determine two additional parameters: corpus size and number of dimensions of the vectors. We set on a 30 million words corpus; as for the dimension of the vectors, we followed (Huckle 1996) and used the 2,000 most common content words in our corpus as dimensions.

Our algorithm for resolving inferential definite descriptions is as follows. For each description, all head nouns and head verbs in the previous five sentences are considered as possible antecedents, as in (Poesio, Vieira, & Teufel 1997). For each antecedent, the distance between the vector associated with the head noun of the inferential description and the vector associated with the possible antecedent is measured; the antecedent whose vector is closest to that of the inferential description is chosen. Three different measures of distance were tried: Manhattan, Euclidean, and Cosine.

We used the British National Corpus[3] for training and the 20 articles from (Poesio, Vieira, & Teufel 1997) to evaluate the results.

Experiments and Results

Experiment 1

In order to get a baseline with respect to which to evaluate the actual performance of the method, we ran an experiment in which the antecedent for each inferential description was chosen randomly. Appropriate[4]

[3] This is a 100-million words collection of both written and spoken language, see `http://info.ox.ac.uk/bnc/`.

[4] An issue to be kept in mind in what follows is that inferential descriptions, unlike other cases of referential expressions, may be related to more than one 'antecedent' in a text, and therefore evaluating the results of a system is more difficult in this case (Poesio & Vieira 1998).

antecedents for 11 out of 203 inferential descriptions–5.4% of the total–were found with this method.

Experiment 2

In this second experiment, we trained and resolved over untagged and lemmatised words. We tried window sizes of 1,2,3,5 and 10 words. The results for the three distance measures were as follows, with the best result in bold:

	Window Size		
Metric	1	2	3
Man	37 (18.2%)	36 (17.7%)	39 (19.2%)
Euc	37 (18.2%)	36 (17.7%)	39 (19.2%)
Cos	39 (19.2%)	36 (17.7%)	39 (19.2%)

	Window Size	
Metric	5	10
Man	41 (20.2%)	37 (18.2%)
Euc	39 (19.2%)	40 (19.7%)
Cos	42 (20.7%)	**45 (22.2%)**

Cosine worked best as a distance measure, and the results were better with bigger windows. The best results for *Manhattan Metric* were achieved at window sizes of three and five; for *Euclidean Distance*, the results seemed to get (slightly) better with larger windows.

Experiment 3

One problem we observed in the second experiment was that lemmatising might create two identical word-forms out of two different lexemes, usually noun and verb, as in *to plan* and *the plan*, and since we did not distinguish between different parts of speech, the algorithm could not tell the difference. In our third experiment we ran the clustering algorithm and the resolution algorithms on texts in which each word had been tagged, so as to avoid the problem encountered in the previous experiment; and we tried larger window sizes, since it appeared from the previous experiment that larger windows performed better.[5] The results are summarised by the following two tables:

	Window size			
Metric	1	2	3	5
Man	34 (16.8%)	35 (17.2%)	41 (20.2%)	41 (20.2%)
Euc	35 (17.2%)	37 (18.2%)	37 (18.2%)	36 (17.7%)
Cos	41 (20.2%)	45 (22.1%)	**46 (22.7%)**	41 (20.2%)

	Window size		
Metric	10	15	20
Man	42 (20.7%)	44 (21.7%)	44 (21.7%)
Euc	37 (18.2%)	38 (18.7%)	39 (19.2%)
Cos	41 (20.2%)	38 (18.7%)	38 (18.7%)

The interesting fact about these results is that although *Cosine* was again the most successful measure when a window size of 3 was used, increasing the window size made things worse, not better; unlike for *Manhattan Metric*, whose performance improved with larger windows. Anyway, the total number of correctly resolved inferential descriptions did not change.

[5]In this second experiment we also tried varying two additional parameters, (i) we ran the clustering algorithm giving equal weight to all words in the window, no matter its distance from the word whose vector was being updated; (ii) we constructed vectors of twice the size, distinguishing between left and right context; but neither of these changes affected the results (Schulte im Walde 1997).

The per-class results for the best-performing combination in the experiments were as follows:

Relationship	Exp. 2		Exp. 3		Total
Same Head	9	(100.0%)	9	(100.0%)	9
Synonymy	3	(25.0%)	4	(33.3%)	12
Hypernymy	2	(15.4%)	2	(15.4%)	13
Meronymy	4	(36.4%)	2	(18.2%)	11
Names	1	(2.3%)	1	(2.3%)	44
Events	3	(10.0%)	5	(16.7%)	30
Compound Nouns	16	(66.7%)	16	(66.7%)	24
Discourse Topic	2	(14.3%)	1	(7.1%)	14
Inference	5	(10.9%)	6	(13.0%)	46
Total	45	(22.2%)	46	(22.7%)	203

Discussion

Analysis of the Results

Even the best parameter configuration (measure *Cosine*, window size of 3) only resulted in appropriate antecedents for 22.7% of the inferential descriptions. Why was that?

The cases in which an inferential description was not resolved to its correct antecedent fell in the following categories:

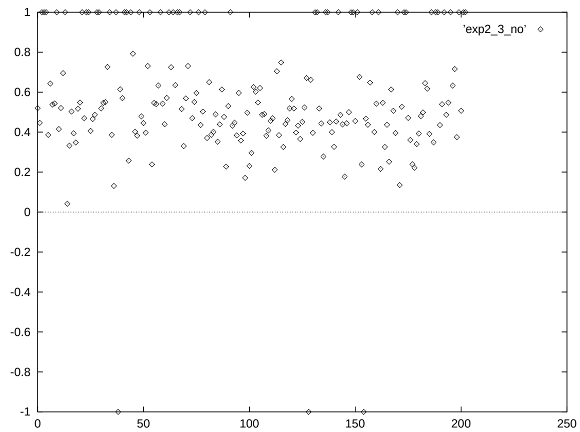
- In some cases, the desired antecedent could not be found since it was not on the list of possible antecedents for the inferential description. This happened if the right word was either before the preceding five sentences, or after the description. In this case, another (incorrect) antecedent was still suggested by the algorithm. There were 22 (10.8%) such cases in the experiments.

- In several cases, there was a word-form among the antecedents which was identical to the inferential description, and therefore always chosen as antecedent, yet it was not the desired antecedent. We already mentioned one reason for that– lemmatisation occasionally created two identical word-forms, e.g., *plan* from *planned*. Another, more interesting reason is that sometimes the desired antecedent was described using a different word-form. This happened, for example, with inferential descriptions referring to names: e.g., one text about companies mentioned the word *company* quite often, and then it mentioned a specific company called *Pinkerton*. The following inferential description *the company* referred to this specific company, but the algorithm picked instead an antecedent explicitly introduced with the word *company* that had appeared in the preceding five sentences.

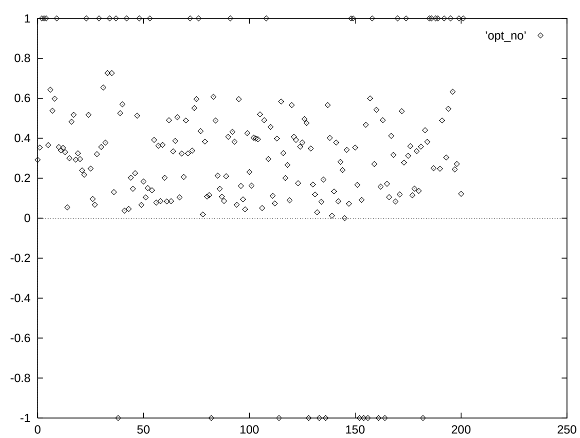
- Finally, the antecedent found by the algorithm was semantically very close to the inferential description– in some cases, even closer–but still not the right antecedent: for example, in one case *market* resolved to *customer* instead of *phone service*. About a third of the problems in both experiments fell in this category.

Semantic Priming and Inferential Descriptions

Even though in both Experiments 2 and 3 we got much better results than chance, and even though the results could still be improved by about 14-15% with better clustering algorithms, the fact that in about a third of the cases the correct antecedent was not the semantically closest one clearly indicates that what we called the Main Hypothesis is false: i.e., that semantic priming is not the only factor involved in the resolution of inferential descriptions. This insight is undermined by the following observation: the following figure shows the cosine of the distances of the 203 antecedents to the inferential descriptions, as chosen by our algorithm. The higher the cosine is (i.e. the closer to +1), the shorter the distance between antecedent and description was. The numbers very between -1 and +1, but are concentrated in the area between 0.3 and 0.6:



The next figure shows the cosine of the distances of the 203 desired antecedents to the inferential descriptions. Also these distances show variation, but in the area between 0 and 0.6:



Even if we had resolved all inferential descriptions to an antecedent in a very short distance, we would not have succeeded, since – as the second figure clearly shows – the resolution should be to an antecedent in a certain distance. The desired antecedent was therefore generally not that word semantically closest to, i.e., most strongly priming, the inferential description.

The most obvious next hypothesis, especially at the light of previous work on definite descriptions, is that attentional mechanisms play a role–i.e., that a focusing mechanism such as those suggested by Grosz (1977) and Sidner (1979) restricts the range of potential antecedents. If this were the case, the 'long-distance' cases of inferential descriptions could then be taken as references to previous discourse foci put on the stack.

Identifying the 'focus' (or 'foci') and tracking focus shifts in a real text in a completely uncontroversial fashion is notoriously difficult, and it is certainly not a task that can be implemented at the moment; we did nevertheless attempt a preliminary verification of this new hypothesis by analysing 4 of the 20 texts previously studied, identifying the available foci according to the proposal in (Poesio, Stevenson, & Hitzeman 1997), and trying to decide for each inferential description whether its resolution only depended on lexical knowledge (i.e., the antecedent was clearly not a focus) or whether instead its antecedent was one of the current foci; we didn't count unclear cases. Surprisingly enough, given all the possible complications just mentioned, the results were fairly clear: of the 44 inferential descriptions in these four texts that we could classify unambiguously, only 15 (about 33%) depended exclusively on lexical knowledge for their resolution; in 29 cases, keeping track of the focus was necessary.

This admittedly very preliminary study suggests that our algorithm in fact performed better than the 22.7% figure mentioned above would suggest. If only about 33% of inferential descriptions can be resolved solely on the ground of lexical knowledge and without keeping track of the current focus, then a fairer evaluation of the performance of our clustering algorithm is that it achieved about 66% of what we could expect it to achieve.

It should also be noted that this analysis indicates that completely ignoring commonsense knowledge during resolution, and just assigning the current focus as antecedent for an inferential description, would not work either: for one thing, about 33% of inferential descriptions do not relate to the current focus, but to some other discourse entity; and anyway when more than one focus is available, the choice among them goes down to lexical knowledge again. In other words, both lexical information and information about the current focus really seem necessary.

Future Work

Since the commonsense knowledge acquired by the methods discussed in this paper does seem to be crucial for resolving inferential descriptions, and the choice of parameters for the clustering algorithm as well as the choice of the measure do seem to have an impact on the results,[6] we intend to continue our investigation of different ways of choosing the dimensions, window sizes, other measures and also combinations of measures, to see if we can improve the method's performance in this respect.

We expect that performance will be improved by using the same corpus for training and evaluation (already, we had to correct for differences between

[6]See (Levy, Bullinaria, & Patel 1997) for a more thorough discussion of the impact of various parameter configurations on different tasks.

British and American lexicon). We are also considering whether more than one type of clustering algorithm may be needed. The particular way of computing similarity we have adopted here looks like a good method for acquiring synonymy relations and subtyping relations, i.e., the information used for resolving descriptions that co-refer with their antecedent without being same-head, such as those descriptions that are expressed via a synonymous or hyponymous predicate (as in *the home / the house*) or that refer to an event (as in *John killed Mary. THE MURDER took place ....*). However, words that are merely associated such as *door / house* do not necessarily always occur in the same contexts; in order to learn this sort of information it may be better to simply look at how often the words themselves occur in the same context, instead of looking at which other words they occur with. E.g., one could tell that 'door' and 'house' are related because they occur often together, especially if they occur together in certain constructions. Vector distance does not expose the desired similarity between 'door' and 'house'; we are investigating the possibility of adding further factors, such as a direct measure of association between the target words, in the decision process. Information from parsing could be useful in the same way.

Acknowledgements

We wish to thank Will Lowe, Scott McDonald and Renata Vieira for much help with the algorithms and with testing the system. Massimo Poesio is supported by an EPSRC Advanced Fellowship.

References

Charniak, E. 1993. *Statistical Language Learning.* Cambridge, MA: The MIT Press.

Church, K. W., and Hanks, P. 1989. Word Association Norms, Mutual Information, and Lexicography. In *Proc. of the 27th Annual Meeting of the Association for Computational Linguistics*, 76–83.

Clark, H. H., and Marshall, C. R. 1981. Definite Reference and Mutual Knowledge. In Joshi, A.; Webber, B.; and Sag, I., eds., *Elements of Discourse Understanding.* New York: Cambridge University Press.

Clark, H. H. 1977. Bridging. In Johnson-Laird, P. N., and Wason, P., eds., *Thinking: Readings in Cognitive Science.* London and New York: Cambridge University Press.

Grosz, B. J., and Sidner, C. L. 1986. Attention, Intention, and the Structure of Discourse. *Computational Linguistics* 12(3):175–204.

Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding.* Ph.D. Dissertation, Stanford University.

Hawkins, J. A. 1978. *Definiteness and Indefiniteness.* London: Croom Helm.

Huckle, C. 1996. *Unsupervised Categorization of Word Meanings Using Statistics and Neural Network Models.* Ph.D. Dissertation, University of Edinburgh.

Levy, J.; Bullinaria, J.; and Patel, M. 1997. Using Co-occurrence Statistics for Modelling Language Processes. In Crocker, M., ed., *Proc. of the AMLAP Conference.*

Lund, K.; Burgess, C.; and Atchley, R. A. 1995. Semantic and Associative Priming in High-Dimensional Semantic Space. In *Proc. of the 17th Annual Conference of the Cognitive Science Society*, 660–665.

Meyer, D. E., and Schvaneveldt, R. 1971. Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations. *Journal of Experimental Psychology* 90:227–235.

Miller, G. A.; Beckwith, R.; Fellbaum, C.; Gross, D.; and Miller, K. J. 1990. Introduction to WordNet: An On-line Lexical Database. *Journal of Lexicography* 3(4).

Poesio, M., and Vieira, R. 1998. A Corpus-Based Investigation of Definite Description Use. *Computational Linguistics.* To appear.

Poesio, M.; Stevenson, R.; and Hitzeman, J. 1997. Global Focus and Pronominalization. In Jurafsky, D., and Regier, T., eds., *Proc. of First Conference on Computational Psycholinguistics.* Berkeley, CA: University of California.

Poesio, M.; Vieira, R.; and Teufel, S. 1997. Resolving Bridging References in Unrestricted Text. In Mitkov, R., ed., *Proc. of the ACL Workshop on Operational Factors in Robust Anaphora Resolution*, 1–6.

Prince, E. F. 1981. Toward a Taxonomy of Given-New Information. In Cole, P., ed., *Radical Pragmatics.* New York: Academic Press. 223–256.

Schütze, H. 1992. Dimensions of Meaning. In *Proc. of Supercomputing*, 787–796.

Schulte im Walde, S. 1997. Resolving Bridging Descriptions in High-Dimensional Space. Studienarbeit, Universities of Stuttgart and Edinburgh.

Sidner, C. L. 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse.* Ph.D. Dissertation, MIT.

Vieira, R., and Poesio, M. 1997. Processing Definite Descriptions in Corpora. In Botley, S., and McEnery, T., eds., *Corpus-based and Computational Approaches to Discourse Anaphora.* UCL Press. To appear.

Vieira, R., and Teufel, S. 1997. Towards Resolution of Bridging Descriptions. In *Proc. of the 35th Joint Meeting of the Association for Computational Linguistics.*

Vieira, R. 1998. *Definite Description Resolution in Unrestricted Texts.* Ph.D. Dissertation, University of Edinburgh, Centre for Cognitive Science.