

Chapter 1

DISTRIBUTIONAL SIMILARITY AND PREPOSITION SEMANTICS

Timothy Baldwin
CSLI, Stanford University
tbaldwin@csli.stanford.edu

Abstract Prepositions are often considered to have too little semantic content or be too polysemous to warrant a proper semantic description. We illustrate the suitability of distributional similarity methods for analysing preposition semantics by way of an inter-preposition similarity task, and make the claim that any semantic account of preposition semantics must be partially conditioned on valence.

Keywords: distributional similarity, latent semantic analysis, preposition semantics, Roget's thesaurus, preposition valence

1. Introduction

While nouns, verbs and adjectives have received considerable attention in terms of both lexical semantic language resource development (Ikehara et al., 1991; Mahesh, 1996; Fellbaum, 1998) and automatic ontology construction (Grefenstette, 1994; Lin, 1998b; Widdows and Dorow, 2002), relatively little work has been done on creating resources for prepositions. Perhaps a large part of the reason for this is that the semantics of a transitive preposition can be bleached and determined largely by the semantics of the head noun it governs (e.g. *at last*, *on Wednesday*, *in question*: (Pustejovsky, 1995)) or its governing verb (e.g. *refer to*, *talk about*). However, many prepositions also have predicative usages (e.g. *time is up*, *the cheese is off*, *flairs are in*), and the semantics of peripheral PPs is determined largely by the preposition (e.g. *from March*, *in Toulouse*, *by the gate*, *at/in Stanford*). Accordingly, some account of preposition semantics seems unavoidable.

There is a relative sparsity of computational research on preposition semantics, which can perhaps be explained by the perception that prepositions are both semantically vacuous and distributionally highly promiscuous, and consequently have a very low information content. This is most pronounced in bag-of-words tasks such as information retrieval where prepositions are generally listed in “stop word” lists for exclusion as index terms.

Our interest is in testing the viability of distributional methods in the derivation of a model of preposition semantics, working under the hypothesis that preposition semantics are stable enough that they can be classified accurately by distributional similarity techniques. Our approach here is based on the distributional hypothesis of (Harris, 1968) that similar words tend to occur in similar linguistic contexts. This observation has been used to explain various aspects of human language processing, from lexical priming (Lund et al., 1995) to retrieval in analogical reasoning (Ramscar and Yarlett, 2003). It has also been employed in a range of natural language processing tasks, including word sense disambiguation (Schütze, 1998) and automatic thesaurus construction (Lin, 1998a). To our knowledge it has not previously been used to analyse the meaning of closed-class words.

As well as demonstrating the ability of similarity methods to capture intuitive correlations in the semantics of prepositions, we are interested in unearthing semantic anomalies between particles and transitive prepositions and motivating a valence-conditioned classification of English prepositions. **Intransitive prepositions** (Huddleston and Pullum, 2002) (which we will interchangeably refer to as **particles**) are valence-saturated and occur most commonly as: (a) components of larger multiword expressions (notably **verb particle constructions**, or VPCs, such as *pick up*, *call in* and *chicken out*), (b) predicates (e.g. *time is **up***, *flairs are **in***) or (c) prenominal modifiers (e.g. *the **up** escalator*, ***off** milk*). **Transitive prepositions**, on the other hand, select for NP complements to form prepositional phrases (PPs, e.g. ***at** home*, ***in** the end*). The bare term **preposition** is valence-underspecified. Hereafter, we will index intransitive prepositions with the suffix “₀” (e.g. *up₀*) and transitive prepositions with the suffix “₁” (e.g. *up₁*) in cases where we wish to refer to a particular valence.

It is relatively easy to find senses which are attested for only intransitive prepositions (e.g. the “hip/in fashion” sense of *in* above) and also uniquely transitive prepositions (e.g. *from*) which by definition do not have intransitive semantics. Of greater interest is the degree of correlation between intransitive and transitive preposition sense according to automatically-derived semantic classifications. That is, we seek to quan-

tify the degree of semantic divergence between intransitive and transitive usages of different prepositions.

One piece of preliminary evidence which underlines the potential applicability of the distributional hypothesis to prepositions comes from the field of English part-of-speech (POS) tagging. All major POS tagsets¹ prefer to underspecify valence (e.g. there is no tag distinction between intransitive and transitive verbs), with the glaring exception of prepositions which are in all cases partitioned into intransitive and transitive instances. If there were a sharp demarcation in wordform between intransitive and transitive prepositions in English, this finding would perhaps not be surprising. However, a summary analysis of the written component of the British National Corpus (BNC, (Burnard, 2000)) reveals that while the type overlap between the two classes is only around 8%, the token overlap is roughly 70%. That is, roughly 70% of preposition token instances are potentially ambiguous between an intransitive and transitive usage. Given that taggers are able to deal with this ambiguity, generally using the immediate lexical context of a given preposition token, it would appear that intransitive and transitive usages of a given preposition are to some degree distributionally dissimilar. In this paper, we seek to confirm that this distributional dissimilarity correlates with semantic disparity, and at the same time determine whether semantically-related prepositions are distributionally similar.

The remainder of this paper is structured as follows. Section 1.2 outlines our implementation of distributional similarity as a means of modelling simplex preposition semantics. Section 1.3 describes the two gold standard sources of English preposition similarity used in this research. Section 1.4 presents quantitative and qualitative evaluation of our method. We outline related research in Section 1.5 and conclude the paper in Section 1.6.

2. Calculating inter-preposition similarity

In this paper, we consider the task of inter-preposition similarity, that is determination of the relative similarity of different preposition pairs. The procedure used to calculate preposition similarity is knowledge-free and based on Latent Semantic Analysis (LSA, (Deerwester et al., 1990)). Our technique is very similar to the approach taken to building a “context space” by (Schütze, 1998). We measured the frequency of co-occurrence of our target words (the 20,000 most frequent words), with a set of 1000 “content-bearing” words (we used the 51st to the 1050th most frequent words, the 50 most frequent being taken to have extremely low information content). A target word was said to co-occur

with a content word if that content word occurred within a window of 5 words to either side of it. In order to overcome data sparseness, we used Singular Value Decomposition (SVD) to reduce the dimensionality of the feature space from 1000 to 100. This limits each target word vector to 100 factors which reflect the patterns of association in the matrix, allowing relations to be discovered between target words even if there is not direct match between their context words. We used the various tools in the GTP software package, created at the University of Tennessee,² to build these matrices from the co-occurrence data and to perform SVD analysis.

The resulting representation is a 100-feature vector for each target word. Using this we can calculate the similarity between two terms by finding the cosine of the angle between their vectors.

As mentioned above, we distinguish prepositions according to valence, and seek to provide evidence for divergences in transitive and intransitive preposition semantics. This is achieved according to Methods PREP_1 and PREP_2 , as detailed below. We evaluate the methods over the written component of the BNC (90m words).

Method PREP_1 First, we ran the above method over wordforms. With this method, we are thus unable to differentiate intransitive and transitive usages of a given preposition.

Method PREP_2 Second, we ran our method including POS tags from the output of the RASP system (Briscoe and Carroll, 2002), i.e. treating each wordform-POS tag pair as a single token. The RASP tagger is based on the CLAWS-4 tagset, and thus offers a fine-grained distinction between different kinds of prepositions and particles. In extracting our context space we collapsed the different varieties of prepositions to give us one category for transitive prepositions and one for intransitive prepositions.

While LSA is generally applied simply to wordforms, we are certainly not the first to integrate POS tags with the wordforms to generate POS-sensitive semantic models. E.g., (Widdows, 2003) demonstrated the superiority of POS-conditioned semantic models on a taxonomy induction task.

3. Gold standard sources of inter-preposition similarity

In order to evaluate the quality of the preposition similarities derived via LSA, we turn to the only two large-scale public-domain resources we are aware of that provide a unified, systematic account of preposition

semantics: the LCS-based preposition lexicon of (Dorr, 1997),³ and the 1911 edition of Roget’s thesaurus.⁴

3.1 LCS-based preposition lexicon

The preposition lexicon of (Dorr, 1997) is couched in lexical conceptual semantics (Jackendoff, 1985), and is made up of 165 English prepositions classified into 122 intransitive and 375 transitive senses. Each preposition sense is described in the form of an LCS-based representation such as (toward Loc (nil 2) (UP Loc (nil 2) (* Thing 6))), corresponding to the “up the stairs” sense of *up*₁. (Resnik and Diab, 2000) propose a method for deriving similarities from LCS representations by: (1) decomposing them into feature sets, (2) calculating the information content $I(f)$ of each unit feature f based on the overall feature distribution, and (3) measuring the similarity between two LCS representations according to:

$$sim_{LCS}(e_1, e_2) = \frac{2 \times I(F(e_1)) \cap I(F(e_2))}{I(F(e_1)) + I(F(e_2))} \quad (1.1)$$

where e_1 and e_2 are lexicon entries, $F(e_i)$ is the decomposed feature set associated with e_i , and $I(F(e_i))$ is the information content of that feature set. (Resnik and Diab, 2000) define the similarity between two words to be the maximum value of $sim_{LCS}(e_1, e_2)$ over the cross product of all lexical entries for the words.

One key feature of this lexicon is that it captures the transitive and intransitive preposition senses separately, but within a common representation. As a result, we are able to derive similarities (a) at the word-form level, comparing all senses of a given preposition pair irrespective of valence, and (b) in a valence-sensitive fashion, calculating sim_{LCS} only for lexicon entries of equivalent valence. This facilitates independent analysis of the correlation of PREP₁ (wordform-based) and PREP₂ (POS-conditioned) with the preposition lexicon-derived similarities.

It is worth pointing out that, in the context of an experiment testing correlation with human judgements on verb similarity, (Resnik and Diab, 2000) found sim_{LCS} to be inferior to a number of taxonomic similarity measures and a distributional similarity measure. It is thus with a certain degree of reservation that we reimplement their method, noting however that the taxonomic similarity avenue is not open to us due to the absence of a taxonomy.

3.2 Roget’s thesaurus

The 1911 edition of Roget’s thesaurus incorporates around 100K lexical entries in a total of 1000 semantic classes. In the original 1911 configuration the classes have no explicit relational structure, although subsequent work has been done to add hierarchical structure to the thesaurus (e.g. (Kirkpatrick, 1988)). We justify our use of the 1911 edition of Roget’s thesaurus on the grounds that (a) there are no restrictions on the use of this version of the thesaurus, and (b) the classification of prepositions is largely unchanged in more recent editions of the thesaurus.

One attraction of Roget’s thesaurus is that, within each semantic class, it lists words according to the four basic word classes of noun, verb, adjective and adverb.⁵ Because of this cross-listing, we can preserve the experimental setup described above for LSA, calculating inter-preposition similarity either according to wordform or conditioned on POS.⁶

In Roget’s thesaurus, prepositions are listed as either adjectives or adverbs, which would superficially appear to correspond to transitive and intransitive prepositions, respectively. In practice, adjectival entries are restricted to predicative and attributive particles such as *the in crowd* and hence limited in number, whereas adverbial entries represent a mix of intransitive and transitive usages. Consider the preposition *up*, for example, which is listed twice as an adjective (BUBBLE, as in *frothy*, and EXCITATION, as in *stung to the quick*⁷) and twice as an adverb (HEIGHT, as in *aloft*, and VERTICALITY, as in *on end*). Here, it is not clear whether the adverbial entries are intended to be intransitive, transitive or both. In some cases, the valence is self-evident as the preposition in question is either uniquely transitive (e.g. *from*, listed under MOTIVE) or uniquely intransitive (e.g. *aback*, listed under REAR). Alternatively, a sense may be particular to a given valence. However, more often than not, we have no reliable way of determining the intended valence of each preposition entry. Thus, we are able to make use of the adjectival entries in modelling particle sense, but have no immediate means of capturing strictly transitive preposition sense. Having said this, for the purposes of evaluation, we consider adjectival preposition entries to be particles and adverbial preposition entries to be transitive prepositions.

Given the lack of hierarchical structure in the 1911 edition of Roget’s thesaurus, our options for deriving class-to-class and word-to-word similarities are restricted. The simplest means of deriving class-to-class similarities is to calculate the relative lexical overlap; word-to-word simi-

larities can equivalently be obtained by calculating the degree of overlap in class membership of each word. Unsurprisingly, this naive methodology suffers from acute data sparseness, culminating in the vast majority of class or word pairings being assigned a similarity of 0. In order to overcome this shortcoming, we notice that it is possible to describe a word pairing by way of a bipartite graph with the classes each word occurs in as the opposing vertices. We can then represent class similarities as edges in the graph, and calculate word-to-word similarity according to the maximal bipartite matching (i.e. set of edges such that every vertex is joined to some other vertex) with the highest mean edge score. We initialise each class similarity $sim_C(i, j)$ to 1 iff $i = j$ and 0 otherwise, such that in the initial configuration, the bipartite graph method is equivalent to the naive class overlap method. We can now iterate between calculating word-to-word and class-to-class similarities—using a bipartite graph with *words* as vertices and *word similarities* as edges in the class-to-class case—and feed the results of the word-to-word similarity recalculations into class-to-class similarity recalculations, and vice versa.

The net effect of this iterative process is to monotonically propagate the effects of class and word overlap, such that both class and word similarities progressively converge to 1. Our driving motivation in this is essentially to “smooth” similarities and eliminate instances of similarity 0. The stopping condition on the method, therefore, is the condition of there being no class similarity $sim_C(i, j)$ or word similarity $sim_W(i, j)$ with value 0. In our experiments, this was generally found to occur on the third iteration.

As we are interested only in preposition similarity (and due to limitations on computational resources), we calculate class-to-class similarities only over those classes which contain one of 54 commonly-occurring prepositions, a total of 78 classes; in calculating word-to-word similarity for non-prepositions contained in the 78 classes, we focus on class membership only over the preposition-containing classes.

In addition to evaluating word-to-word preposition similarity according to simplex preposition entries, we test the use of VPCs as a proxy to situated particle semantics. The method here is identical to that for simplex words, except that we additionally look for occurrences of VPCs as contained in a list of VPC types extracted out of the BNC (Baldwin and Villavicencio, 2002), and record each such occurrence as an instance of the particle contained therein. That is, we do not distinguish between simplex occurrences of the preposition and occurrences within VPCs. This is not intended to be a general claim about semantic headedness or the relative semantic contribution of the particle in VPCs. Rather

PREP ₁	PREP ₂	
(A) all	(B) intransitive	(C) transitive
0.304	0.365	0.386

Table 1.1. Correlation (r) between the LCS-based and LSA similarities

we are testing the hypothesis that particles with similar semantics will occur with the same classes of verbs.

Due to the inherent complexity of the similarity calculation, we restrict the number of VPCs by counting the VPCs contained in each class not containing a simplex preposition, and including only those VPCs found in the 200 most heavily VPC-populated classes.

4. Evaluation

In this section, we evaluate the LSA-based similarities relative to similarities derived from the LCS lexicon and also Roget’s thesaurus.

We measure the correlation between the distributional similarities and both LCS- and thesaurus-derived similarities according to Pearson’s r , as applied to the attested pairings of the nine prepositions *about*, *down*, *in*, *off*, *on*, *out*, *over*, *through* and *up*. We determine the correlation for three distinct datasets: (A) preposition similarity according to PREP₁ (with underspecification of valence); (B) particle similarity according to PREP₂; and (C) transitive preposition similarity according to PREP₂. In the case of (A), therefore, we calculate the distributional similarity of prepositions in the absence of POS information, and likewise do not distinguish between intransitive and transitive prepositions in the LCS lexicon. For (B) and (C), on the other hand, we consider only prepositions of fixed transitivity in both the BNC data and LCS lexicon.

4.1 Correlation with LCS-based similarities

The mean r values relative to the LCS-based similarities are given in Table 1.1 for datasets A, B and C. While the values are relatively modest, they provide weak evidence for the ability of LSA to capture preposition semantics. Perhaps more importantly, the correlations for the intransitive and transitive preposition similarity tasks ((B) and (C), respectively) are higher than that for the valence-underspecified preposition similarity task (A), at a level of statistical significance (based on the two-tailed t -test, $p < .05$). This suggests that our model of preposition semantics is more stable when valence is specified, providing tentative

ROGET'S SIMILARITY					
	−VALENCE	−VALENCE _{VPC}	+VALENCE	+VALENCE _{VPC}	VPC
PREP ₁ (A)	0.004	0.080	—	—	0.097
PREP ₂ (B)	−0.183	0.881	−0.235	0.863	0.805
PREP ₂ (C)	−0.173	−0.287	−0.258	−0.205	−0.197

Table 1.2. Correlation (r) between the Roget's thesaurus-based and LSA similarities

support for the claim that preposition semantics are to some degree conditioned on valence.

Recall that we had reservations about the quality of similarities produced with this method, based on the findings of (Resnik and Diab, 2000) over a small-scale verb similarity task. Having said this, the fact that both valence-specified models of distributional similarity were found to correlate more highly than the valence-underspecified model would appear to be significant.

4.2 Correlation with Roget's-based similarities

We turn next to Roget's thesaurus and calculate the correlation with similarities derived: (a) independently of valence information for simplex preposition entries (conflating adjectival and adverbial preposition entries: −VALENCE), optionally incorporating semantic classes for VPCs (−VALENCE_{VPC}); (b) conditioned on valence information (+VALENCE), once again optionally incorporating semantic classes for VPCs (+VALENCE_{VPC}); or (c) based only on the VPC entries, without the simplex preposition entries (VPC). The results are presented in Table 1.2. Note that we compare PREP₁ against only the valence-underspecified Roget's similarities as there is no obvious way of combining similarities across the two transitivityes for a given preposition. Note also that for the valence-specified models, we always compare like with like, e.g. PREP₂ (B) is only compared against particle similarities in the correlation analysis with +VALENCE and +VALENCE_{VPC}.

We see some interesting results. First, the correlation for PREP₁ is almost 0 in all cases. That is, in the absence of valence information, the relative similarity values for the different prepositions are nearly randomly distributed. Next, the transitive preposition similarities (the row of PREP₂ (C)) are negatively correlated in all cases, but relatively low. Thorough error analysis is required to determine the cause of this negative correlation, but it is worth noting the differential between the r values for PREP₂ (C) and PREP₁ down each column, indicating that the LSA

similarities for valence-underspecified prepositions vary significantly over those for transitive prepositions. Recall that, for the purposes of this evaluation, we are treating adverbial preposition entries in Roget’s as transitive prepositions, and it is the similarities for these that we are comparing PREP_2 (C) against. Given our observations above about the mixed nature of adverbial prepositions, it is perhaps not surprising that no real correlation was found. Finally, PREP_2 (B) produces remarkably similar results to PREP_2 (C) for the models which do not make use of the VPC data, but when we add in the VPC classes, we find the correlation to be surprisingly high. The combination of VPC data and valence-underspecified preposition entries returns the highest r value at 0.881. This compares very favourably with the $r = 0.901$ and $r = 0.793$ figures cited as inter-annotator correlation for noun and verb similarity tasks (Resnik, 1995; Resnik and Diab, 2000). Indeed, it provides strong evidence that, at least when viewed in the context of VPC occurrence, particle semantics are well-defined and can be captured effectively by distributional similarity methods.

4.3 Discussion of the results

What the above experiments show is, first, that LSA can be applied successfully to the task of inter-preposition similarity modelling. This in itself is a surprising finding, given that the standard practice in established domains for LSA such as information retrieval (IR) is to ignore all prepositions and other stop words. This result is particularly striking as it was validated over heterogeneous sets of similarities, derived from formal semantic representations in the first instance and word clusters in the second.

In our second experiment based on Roget’s thesaurus, we found that complementing the simplex inventory of preposition sense led to a huge increase in correlation with the LSA similarities. One could possibly argue that this finding is a by-product of the fact that we are deriving our similarities from Roget’s in a similar fashion to LSA, in that we are making use of a context window in calculating the similarities. However, when we consider what role the VPCs are playing in the similarity calculation, it quickly becomes evident that this is not the case. At no point do we compare which verbs different particles co-occur with. Instead, we take note of which semantic classes those VPCs occur in, and base our similarity calculation on class overlap as per usual. That is, for two particles to be similar, they must combine with verbs (and not necessarily the same verbs) to generate VPCs of the same semantic types. As an illustration of this, consider the particle pairing *back₀* and

*down*₀. In the +VALENCE model, $\text{sim}_W(\text{back}_0, \text{down}_0) = 0$ as *down* is not listed as an adjective in Roget's. In +VALENCE_{VPC}, on the other hand, $\text{sim}_W(\text{back}_0, \text{down}_0) = 0.68$ as VPCs such as *fall back/back down* and *keep back/tie down* occur in the same semantic classes.

In the first experiment based on the LCS lexicon, we were able to demonstrate modest gains in correlation by conditioning similarity on valence for both transitive and intransitive prepositions. In the second experiment based on Roget's thesaurus, on the other hand, we provided conclusive evidence that LSA is more adept at capturing particle semantics than the semantics of valence-underspecified prepositions. Taken together, these provide solid evidence that LSA produces higher-quality results in the presence of valence information. We attribute this to semantic disparities between intransitive and transitive forms of a given preposition, or to think of it in set terms, the semantics of each of the two transitivities constitutes a proper subset of that the (valence-underspecified) whole.

Due to the nature of Roget's thesaurus, we were unable to furnish evidence for the stability of transitive preposition semantics in the second experiment. The determination of alternate methods for deriving the semantics of transitive prepositions is left as an item for future research.

5. Related research

Past computational research on preposition semantics falls into two basic categories: large-scale symbolic accounts of preposition semantics, and disambiguation of PP sense. (Cannesson and Saint-Dizier, 2002) developed an LCS-based formal description of the semantics of 170 French prepositions in a similar vein to (Dorr, 1997), but paying particular attention to their corpus usage. (Litkowski, 2002) used digraph analysis to induce a preposition hierarchy, based upon which he proposed disambiguation rules to map preposition sense onto the hierarchy. (O'Hara and Wiebe, 2003) focused exclusively on the disambiguation task, classifying PP tokens according to their case-role in the style of the Penn treebank.

There is also a small body of computational research on prepositions in the context of verb particle constructions. Notably, (Bannard et al., 2003) used distributional similarity between VPCs and their component verbs and prepositions to predict whether the semantics of the simplex words were preserved in VPC; indeed, the LSA similarities used herein derive from this earlier work. Similarly, (McCarthy et al., 2003) and (Baldwin et al., 2003) tested distributional similarity in various forms as a means of predicting the relative compositionality of a given VPC.

6. Conclusion

We have illustrated how distributional similarity methods can be used to successfully calculate inter-preposition similarity, and provided evidence for the valence-dependence of preposition semantics. More generally, we have furnished counter-evidence to the claim that prepositions are ill-suited to distributional similarity methods, in the form of the inter-preposition similarity task. Our hope is that this research will open the way to research on automatically-derived preposition thesauri to act as the catalyst in the development of preposition ontologies.

There is scope for this research to be extended in the direction of empirically-grounded evaluation of inter-preposition similarity, perhaps using human judgements. We are also interested in the impact of dependency data on the semantic classification of prepositions. These are left as items for future research.

Acknowledgements

This research would not have been possible without the assistance of Colin Bannard, who provided the LSA similarities and provided valuable input at various points throughout the development of this research. I would also like to thank John Beavers, Aline Villavicencio and the two anonymous reviewers for their valuable input on this research. The author is supported by the National Science Foundation under Grant No. BCS-0094638 and also the Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University.

Notes

1. By which we specifically refer to the International Corpus of English, Penn and various CLAWS tagsets.
2. <http://www.cs.utk.edu/~lsi/soft.html>
3. <http://www.umiacs.umd.edu/~bonnie/AZ-preps-English.lcs>
4. As distributed by Project Gutenberg: <http://www.gutenberg.net/etext91/roget15a.txt>
5. Interjections and phrases are also optionally listed.
6. Note that this would not be possible in WordNet, e.g., as adjectives and adverbs are listed in independent ontologies.
7. Both of which are antiquated usages.

References

- Baldwin, Timothy, Bannard, Colin, Tanaka, Takaaki, and Widdows, Dominic (2003). An empirical model of multiword expression decomposability. In *Proc. of the ACL-2003 Workshop on Multiword Expression*

- sions: Analysis, Acquisition and Treatment*, pages 89–96, Sapporo, Japan.
- Baldwin, Timothy and Villavicencio, Aline (2002). Extracting the unextractable: A case study on verb-particles. In *Proc. of the 6th Conference on Natural Language Learning (CoNLL-2002)*, pages 98–104, Taipei, Taiwan.
- Bannard, Colin, Baldwin, Timothy, and Lascarides, Alex (2003). A statistical approach to the semantics of verb-particles. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 65–72, Sapporo, Japan.
- Briscoe, Ted and Carroll, John (2002). Robust accurate statistical annotation of general text. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1499–1504, Las Palmas, Canary Islands.
- Burnard, Lou (2000). *User Reference Guide for the British National Corpus*. Technical report, Oxford University Computing Services.
- Cannesson, Emmanuelle and Saint-Dizier, Patrick (2002). Defining and representing preposition senses: A preliminary analysis. In *Proc. of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pages 25–31, Philadelphia, USA.
- Deerwester, Scott, Dumais, Susan, Furnas, George, Landauer, Thomas, and Harshman, Richard (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.
- Dorr, Bonnie J. (1997). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):271–322.
- Fellbaum, Christiane, editor (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.
- Grefenstette, Gregory (1994). *Explorations in Automatic Thesaurus Extractions*. Kluwer Academic Publishers.
- Harris, Zellig (1968). *Mathematical Structures of Language*. Wiley, New York, USA.
- Huddleston, Rodney and Pullum, Geoffrey K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK.
- Ikehara, Satoru, Shirai, Satoshi, Yokoo, Akio, and Nakaiwa, Hiromi (1991). Toward an MT system without pre-editing – effects of new methods in **ALT-J/E**-. In *Proc. of the Third Machine Translation Summit (MT Summit III)*, pages 101–106, Washington DC, USA.

- Jackendoff, Ray (1985). Semantic structure and conceptual structure. In *Semantics and Cognition*, chapter 1, pages 3–22. MIT Press, Cambridge, USA.
- Kirkpatrick, Betty (1988). *Roget's Thesaurus of English Words and Phrases*. Penguin, Harmondsworth, England.
- Lin, Dekang (1998a). Automatic retrieval and clustering of similar words. In *Proc. of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics (COLING/ACL-98)*, Montreal, Canada.
- Lin, Dekang (1998b). Extracting collocations from text corpora. In *Proc. of the COLING-ACL'98 Workshop on Computational Terminology*, Montreal, Canada.
- Litkowski, Kenneth C. (2002). Digraph analysis of dictionary preposition definitions. In *Proc. of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pages 9–16, Philadelphia, USA.
- Lund, Kevin, Burgess, Curt, and Atchley, Ruth Ann (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 660–5, Pittsburgh, USA.
- Mahesh, Kavi (1996). *Ontology Development for Machine Translation: Ideology and Methodology*. Technical Report MCCS-96-292, Computing Research Laboratory, NMSU.
- McCarthy, Diana, Keller, Bill, and Carroll, John (2003). Detecting a continuum of compositionality in phrasal verbs. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.
- O'Hara, Tom and Wiebe, Janyce (2003). Preposition semantic classification via Treebank and FrameNet. In *Proc. of the 7th Conference on Natural Language Learning (CoNLL-2003)*, pages 79–86, Edmonton, Canada.
- Pustejovsky, James (1995). *The Generative Lexicon*. MIT Press, Cambridge, USA.
- Ramscar, Michael and Yarlett, Dan (2003). Semantic grounding in models of analogy: An environmental approach. *Cognitive Science*, (27):41–71.
- Resnik, Philip (1995). Using information content to evaluate semantic similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*.
- Resnik, Philip and Diab, Mona (2000). Measuring verb similarity. In *Proc. of the 22nd Annual Meeting of the Cognitive Science Society (CogSci 2000)*, pages 399–404.

- Schütze, Hinrich (1998). Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- Widdows, Dominic (2003). Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In *Proc. of the 3rd International Conference on Human Language Technology Research and 4th Annual Meeting of the NAACL (HLT-NAACL 2003)*, pages 276–83, Edmonton, Canada.
- Widdows, Dominic and Dorow, Beate (2002). A graph model for unsupervised lexical acquisition. In *Proc. of the 19th International Conference on Computational Linguistics (COLING 2002)*, pages 1093–9, Taipei, Taiwan.