Modelling the substitutability of discourse connectives

Ben Hutchinson
School of Informatics
University of Edinburgh
B.Hutchinson@sms.ed.ac.uk

Abstract

Processing discourse connectives is important for tasks such as discourse parsing and generation. For these tasks, it is useful to know which connectives can signal the same coherence relations. This paper presents experiments into modelling the substitutability of discourse connectives. It shows that substitutability effects distributional similarity. A novel variance-based function for comparing probability distributions is found to assist in predicting substitutability.

1 Introduction

Discourse coherence relations contribute to the meaning of texts, by specifying the relationships between semantic objects such as events and propositions. They also assist in the interpretation of anaphora, verb phrase ellipsis and lexical ambiguities (Hobbs, 1985; Kehler, 2002; Asher and Lascarides, 2003). Coherence relations can be implicit, or they can be signalled explicitly through the use of discourse connectives, e.g. *because, even though*.

For a machine to interpret a text, it is important that it recognises coherence relations, and so as explicit markers discourse connectives are of great assistance (Marcu, 2000). When discourse connectives are not present, the task is more difficult. For such cases, unsupervised approaches have been developed for predicting relations, by using sentences containing discourse connectives as training data (Marcu and Echihabi, 2002; Lapata and Lascarides, 2004). However the nature of the relationship between the coherence relations signalled by discourse connectives and their empirical distributions has to date been poorly understood. In particular, one might wonder whether connectives with similar meanings also have similar distributions.

Concerning natural language generation, texts are easier for humans to understand if they are coherently structured. Addressing this, a body of research has considered the problems of generating appropriate discourse connectives (for example (Moser and Moore, 1995; Grote and Stede, 1998)). One such problem involves choosing which connective to generate, as the mapping between connectives and relations is not one-to-one, but rather many-to-many. Siddharthan (2003) considers the task of paraphrasing a text while preserving its rhetorical relations. Clauses conjoined by *but*, *or* and *when* are separated to form distinct orthographic sentences, and these conjunctions are replaced by the discourse adverbials *however*, *otherwise* and *then*, respectively.

The idea underlying Siddharthan's work is that one connective can be substituted for another while preserving the meaning of a text. Knott (1996) studies the substitutability of discourse connectives, and proposes that substitutability can motivate theories of discourse coherence. Knott uses an empirical methodology to determine the substitutability of pairs of connectives. However this methodology is manually intensive, and Knott derives relationships for only about 18% of pairs of connectives. It would thus be useful if substitutability could be predicted automatically.

Proceedings of the 43rd Annual Meeting of the ACL, pages 149–156,
Ann Arbor, June 2005.

This paper proposes that substitutability can be predicted through statistical analysis of the contexts in which connectives appear. Similar methods have been developed for predicting the similarity of nouns and verbs on the basis of their distributional similarity, and many distributional similarity functions have been proposed for these tasks (Lee, 1999). However substitutability is a more complex notion than similarity, and we propose a novel variance-based function for assisting in this task.

This paper constitutes a first step towards predicting substitutability of cnonectives automatically. We demonstrate that the substitutability of connectives has significant effects on both distributional similarity and the new variance-based function. We then attempt to predict substitutability of connectives using a simplified task that factors out the prior likelihood of being substitutable.

2 Relationships between connectives

Two types of relationships between connectives are of interest: similarity and substitutability.

2.1 Similarity

The concept of lexical similarity occupies an important role in psychology, artificial intelligence, and computational linguistics. For example, in psychology, Miller and Charles (1991) report that psychologists 'have largely abandoned "synonymy" in favour of "semantic similarity".' In addition, work in automatic lexical acquisition is based on the proposition that distributional similarity correlates with semantic similarity (Grefenstette, 1994; Curran and Moens, 2002; Weeds and Weir, 2003).

Several studies have found subjects' judgements of semantic similarity to be robust. For example, Miller and Charles (1991) elicit similarity judgements for 30 pairs of nouns such as *cord–smile*, and found a high correlation with judgements of the same data obtained over 25 years previously (Rubenstein and Goodenough, 1965). Resnik (1999) repeated the experiment, and calculated an inter-rater agreement of 0.90. Resnik and Diab (2000) also performed a similar experiment with pairs of verbs (e.g. *bathe–kneel*). The level of inter-rater agreement was again significant ($r = 0.76$).

1. Take an instance of a discourse connective in a corpus. Imagine you are the writer that produced this text, but that you need to choose an alternative connective.
2. Remove the connective from the text, and insert another connective in its place.
3. If the new connective achieves the same discourse goals as the original one, it is considered **substitutable** in this context.

Figure 1: Knott's Test for Substitutability

Given two words, it has been suggested that if words have the similar meanings, then they can be expected to have similar contextual distributions. The studies listed above have also found evidence that similarity ratings correlate positively with the distributional similarity of the lexical items.

2.2 Substitutability

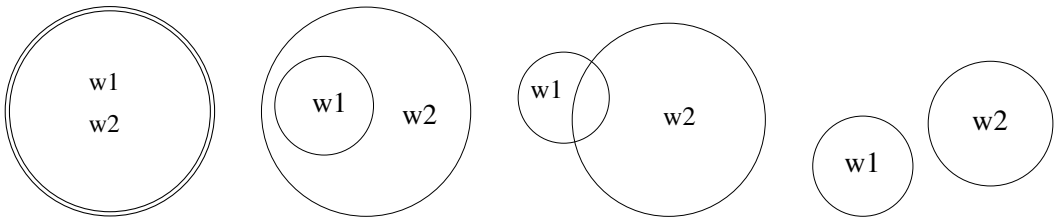
The notion of substitutability has played an important role in theories of lexical relations. A definition of synonymy attributed to Leibniz states that two words are synonyms if one word can be used in place of the other without affecting truth conditions.

Unlike similarity, the substitutability of discourse connectives has been previously studied. Halliday and Hasan (1976) note that in certain contexts *otherwise* can be paraphrased by *if not*, as in

(1) It's the way I like to go to work.
One person and one line of enquiry at a time.
Otherwise/if not, there's a muddle.

They also suggest some other extended paraphrases of *otherwise*, such as *under other circumstances*.

Knott (1996) systematises the study of the substitutability of discourse connectives. His first step is to propose a Test for Substitutability for connectives, which is summarised in Figure 1. An application of the Test is illustrated by (2). Here *seeing as* was the connective originally used by the writer, however *because* can be used instead.



(a) w_1 and w_2 are SYNONYMS

(b) w_1 is a HYPONYM of w_2

(c) w_1 and w_2 are CONTINGENTLY SUBSTITUTABLE

(d) w_1 and w_2 are EXCLUSIVE

Figure 2: Venn diagrams representing relationships between distributions

(2) **Seeing as/because** we've got nothing but circumstantial evidence, it's going to be difficult to get a conviction. (Knott, p. 177)

However the ability to substitute is sensitive to the context. In other contexts, for example (3), the substitution of *because* for *seeing as* is not valid.

(3) It's a fairly good piece of work, **seeing as/#because** you have been under a lot of pressure recently. (Knott, p. 177)

Similarly, there are contexts in which *because* can be used, but *seeing as* cannot be substituted for it:

(4) That proposal is useful, **because/#seeing as** it gives us a fallback position if the negotiations collapse. (Knott, p. 177)

Knott's next step is to generalise over all contexts a connective appears in, and to define four substitutability relationships that can hold between a pair of connectives w_1 and w_2. These relationships are illustrated graphically through the use of Venn diagrams in Figure 2, and defined below.

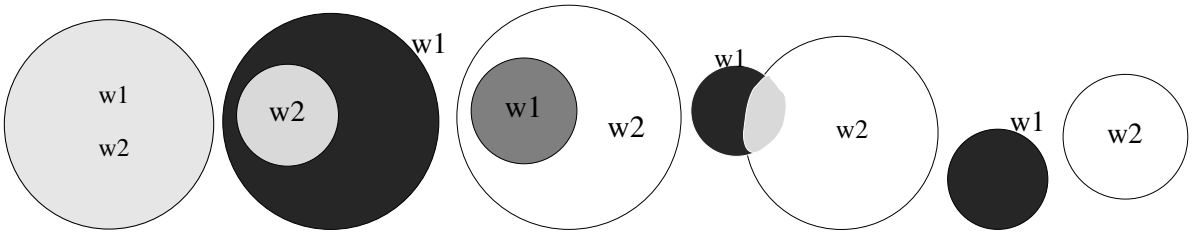
- w_1 is a SYNONYM of w_2 if w_1 can always be substituted for w_2, and vice versa.
- w_1 and w_2 are EXCLUSIVE if neither can ever be substituted for the other.
- w_1 is a HYPONYM of w_2 if w_2 can always be substituted for w_1, but not vice versa.
- w_1 and w_2 are CONTINGENTLY SUBSTITUTABLE if each can sometimes, but not always, be substituted for the other.

Given examples (2)–(4) we can conclude that *because* and *seeing as* are CONTINGENTLY SUBSTITUTABLE (henceforth "CONT. SUBS."). However this is the only relationship that can be established using a finite number of linguistic examples. The other relationships all involve generalisations over all contexts, and so rely to some degree on the judgement of the analyst. Examples of each relationship given by Knott (1996) include: *given that* and *seeing as* are SYNONYMS, *on the grounds that* is a HYPONYM of *because*, and *because* and *now that* are EXCLUSIVE.

Although substitutability is inherently a more complex notion than similarity, distributional similarity is expected to be of some use in predicting substitutability relationships. For example, if two discourse connectives are SYNONYMS then we would expect them to have similar distributions. On the other hand, if two connectives are EXCLUSIVE, then we would expect them to have dissimilar distributions. However if the relationship between two connectives is HYPONYMY or CONT. SUBS. then we expect to have partial overlap between their distributions (consider Figure 2), and so distributional similarity might not distinguish these relationships.

The Kullback-Leibler (KL) divergence function is a distributional similarity function that is of particular relevance here since it can be described informally in terms of substitutability. Given co-occurrence distributions p and q, its mathematical definition can be written as:

$$D(p||q) = \sum_x p(x)(\log \frac{1}{q(x)} - \log \frac{1}{p(x)}) \quad (5)$$



(a) w_1 and w_2 are SYNONYMS (b) w_2 is a HYPONYM of w_1 (c) w_1 is a HYPONYM of w_2 (d) w_1 and w_2 are CONT. SUBS. (e) w_1 and w_2 are EXCLUSIVE

Figure 3: Surprise in substituting w_2 for w_1 (darker shading indicates higher surprise)

The value $\log \frac{1}{p(x)}$ has an informal interpretation as a measure of how surprised an observer would be to see event x, given prior likelihood expectations defined by p. Thus, if p and q are the distributions of words w_1 and w_2 then

$$D(p||q) = E_p(\text{surprise in seeing } w_2 - \text{surprise in seeing } w_1) \quad (6)$$

where E_p is the expectation function over the distribution of w_1 (i.e. p). That is, KL divergence measures how much more surprised we would be, on average, to see word w_2 rather than w_1, where the averaging is weighted by the distribution of w_1.

3 A variance-based function for distributional analysis

A distributional similarity function provides only a one-dimensional comparison of two distributions, namely how similar they are. However we can obtain an additional perspective by using a variance-based function. We now introduce a new function V by taking the variance of the surprise in seeing w_2, over the contexts in which w_1 appears:

$$V(p,q) = Var(\text{surprise in seeing } w_2) = E_p((E_p(\log \frac{1}{q(x)}) - \log \frac{1}{q(x)})^2) \quad (7)$$

Note that like KL divergence, $V(p,q)$ is asymmetric.

We now consider how the substitutability of connectives affects our expectations of the value of V. If two connectives are SYNONYMS then each can always be used in place of other. Thus we would always expect a low level of surprise in seeing one connective in place of the other, and this low level of surprise is indicated via light shading in Figure 3a. It follows that the variance in surprise is low. On the other hand, if two connectives are EXCLUSIVE then there would always be a high degree of surprise in seeing one in place of the other. This is indicated using dark shading in Figure 3e. Only one set is shaded because we need only consider the contexts in which w_1 is appropriate. In this case, the variance in surprise is again low. The situation is more interesting when we consider two connectives that are CONT. SUBS.. In this case substitutability (and hence surprise) is dependent on the context. This is illustrated using light and dark shading in Figure 3d. As a result, the variance in surprise is high. Finally, with HYPONYMY, the variance in surprise depends on whether the original connective was the HYPONYM or the HYPERNYM.

Relationship of w_1 to w_2	Function			
	$D(p\|\|q)$	$D(q\|\|p)$	$V(p,q)$	$V(q,p)$
SYNONYM	Low	Low	Low	Low
HYPONYM	Low	Medium	Low	High
CONT. SUBS.	Medium	Medium	High	High
EXCLUSIVE	High	High	Low	Low

Table 1: Expectations for distributional functions

Table 1 summarises our expectations of the values of KL divergence and V, for the various substitutability relationships. (KL divergence, unlike most similarity functions, is sensitive to the order of arguments related by hyponymy (Lee, 1999).) The

Something happened **and** *something else happened.*
Something happened **or** *something else happened.*
◯ 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5

Figure 4: Example experimental item

	Mean	HYP	CONT. SUBS.	EXCL
SYNONYM	3.97	*	*	*
HYPONYM	3.43		*	*
CONT. SUBS.	1.79			*
EXCLUSIVE	1.08			

Table 2: Similarity by substitutability relationship

experiments described below test these expectations using empirical data.

4 Experiments

We now describe our empirical experiments which investigate the connections between a) subjects' ratings of the similarity of discourse connectives, b) the substitutability of discourse connectives, and c) KL divergence and the new function V applied to the distributions of connectives. Our motivation is to explore how distributional properties of words might be used to predict substitutability. The experiments are restricted to connectives which relate clauses within a sentence. These include coordinating conjunctions (e.g. *but*) and a range of subordinators including conjunctions (e.g. *because*) as well as phrases introducing adverbial clauses (e.g. *now that, given that, for the reason that*). Adverbial discourse connectives are therefore not considered.

4.1 Experiment 1: Subject ratings of similarity

This experiment tests the hypotheses that 1) subjects agree on the degree of similarity between pairs of discourse connectives, and 2) similarity ratings correlate with the degree of substitutability.

4.1.1 Methodology

We randomly selected 48 pairs of discourse connectives such that there were 12 pairs standing in each of the four substitutability relationships.To do this, we used substitutability judgements made by Knott (1996), supplemented with some judgements of our own. Each experimental item consisted of the two discourse connectives along with dummy clauses, as illustrated in Figure 4. The format of the experimental items was designed to indicate how a phrase could be used as a discourse connective (e.g. it may not be obvious to a subject that the phrase *the moment* is a discourse connective), but without providing complete semantics for the clauses, which might bias the subjects' ratings. Forty native speakers of English participated in the experiment, which was conducted remotely via the internet.

4.1.2 Results

Leave-one-out resampling was used to compare each subject's ratings are with the means of their peers' (Weiss and Kulikowski, 1991). The average inter-subject correlation was 0.75 (Min = 0.49, Max = 0.86, StdDev = 0.09), which is comparable to previous results on verb similarity ratings (Resnik and Diab, 2000). The effect of substitutability on similarity ratings can be seen in Table 2. Post-hoc Tukey tests revealed all differences between means in Table 2 to be significant.

The results demonstrate that subjects' ratings of connective similarity show significant agreement and are robust enough for effects of substitutability to be found.

4.2 Experiment 2: Modelling similarity

This experiment compares subjects' ratings of similarity with lexical co-occurrence data. It hypothesises that similarity ratings correlate with distributional similarity, but that neither correlates with the new variance in surprise function.

4.2.1 Methodology

Sentences containing discourse connectives were gathered from the British National Corpus and the world wide web, with discourse connectives identified on the basis of their syntactic contexts (for details, see Hutchinson (2004b)). The mean number of sentences per connective was about $32,000$, although about 12% of these are estimated to be errors. From these sentences, lexical co-occurrence data were collected. Only co-occurrences with dis-

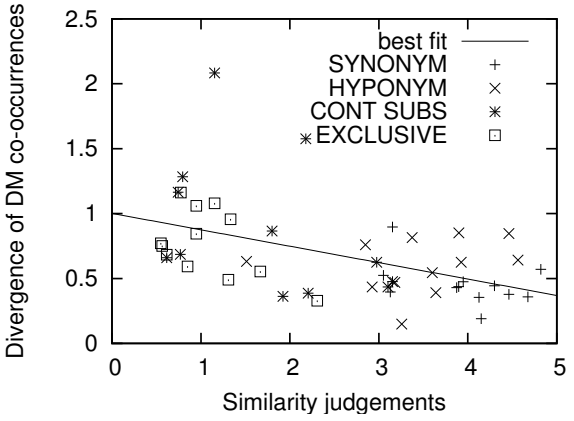


Figure 5: Similarity versus distributional divergence

course adverbials and other structural discourse connectives were stored, as these had previously been found to be useful for predicting semantic features of connectives (Hutchinson, 2004a).

4.2.2 Results

A skewed variant of the Kullback-Leibler divergence function was used to compare co-occurrence distributions (Lee, 1999, with $\alpha = 0.95$). Spearman's correlation coefficient for ranked data showed a significant correlation ($r = -0.51$, $p < 0.001$). (The correlation is negative because KL divergence is lower when distributions are more similar.) The strength of this correlation is comparable with similar results achieved for verbs (Resnik and Diab, 2000), but not as great as has been observed for nouns (McDonald, 2000). Figure 5 plots the mean similarity judgements against the distributional divergence obtained using discourse markers, and also indicates the substitutability relationship for each item. (Two outliers can be observed in the upper left corner; these were excluded from the calculations.)

The "variance in surprise" function introduced in the previous section was applied to the same co-occurrence data.[1] These variances were compared to distributional divergence and the subjects' similarity ratings, but in both cases Spearman's correlation coefficient was not significant.

In combination with the previous experiment, these results demonstrate a three way correspondence between the human ratings of the similarity of a pair of connectives, their substitutability relationship, and their distributional similarity. Hutchinson (2005) presents further experiments on modelling connective similarity, and discusses their implications. This experiment also provides empirical evidence that the new variance in surprise function is not a measure of similarity.

4.3 Experiment 3: Predicting substitutability

The previous experiments provide hope that substitutability of connectives might be predicted on the basis of their empirical distributions. However one complicating factor is that EXCLUSIVE is by far the most likely relationship, holding between about 70% of pairs. Preliminary experiments showed that the empirical evidence for other relationships was not strong enough to overcome this prior bias. We therefore attempted two pseudodisambiguation tasks which eliminated the effects of prior likelihoods. The first task involved distinguishing between the relationships whose connectives subjects rated as most similar, namely SYNONYMY and HYPONYMY. Triples of connectives $\langle p, q, q' \rangle$ were collected such that SYNONYM(p, q) and either HYPONYM(p, q') or HYPONYM(q', p) (we were not attempting to predict the order of HYPONYMY). The task was then to decide automatically which of q and q' is the SYNONYM of p.

The second task was identical in nature to the first, however here the relationship between p and q was either SYNONYMY or HYPONYMY, while p and q' were either CONT. SUBS. or EXCLUSIVE. These two sets of relationships are those corresponding to high and low similarity, respectively. In combination, the two tasks are equivalent to predicting SYNONYMY or HYPONYMY from the set of all four relationships, by first distinguishing the high similarity relationships from the other two, and then making a finer-grained distinction between the two.

4.3.1 Methodology

Substitutability relationships between 49 structural discourse connectives were extracted from Knott's (1996) classification. In order to obtain more evaluation data, we used Knott's methodology to obtain relationships between an additional 32 connec-

[1] In practice, the skewed variant $V(p, 0.95q + 0.05p)$ was used, in order to avoid problems arising when $q(x) = 0$.

	$max(D_1, D_2)$	$max(V_1, V_2)$	$(V_1 - V_2)^2$
SYN	0.627	4.44	3.29
HYP	0.720	5.16	8.02
CONT	1.057	4.85	7.81
EXCL	1.069	4.79	7.27

Table 3: Distributional analysis by substitutability

Input to Gaussian Model	SYN vs HYP	SYN/HYP vs EX/CONT
$max(D_1, D_2)$	50.0%	76.1%
$max(V_1, V_2)$	84.8%	60.6%

Table 4: Accuracy on pseudodisambiguation task

tives. This resulted in 46 triples $\langle p, q, q' \rangle$ for the first task, and 10,912 triples for the second task.

The co-occurrence data from the previous section were re-used. These were used to calculate $D(p||q)$ and $V(p, q)$. Both of these are asymmetric, so for our purposes we took the maximum of applying their arguments in both orders. Recall from Table 1 that when two connectives are in a HYPONYMY relation we expect V to be sensitive to the order in which the connectives are given as arguments. To test this, we also calculated $(V(p, q) - V(q, p))^2$, i.e. the square of the difference of applying the arguments to V in both orders. The average values are summarised in Table 3, with D_1 and D_2 (and V_1 and V_2) denoting different orderings of the arguments to D (and V), and max denoting the function which selects the larger of two numbers.

These statistics show that our theoretically motivated expectations are supported. In particular, (1) SYNONYMOUS connectives have the least distributional divergence and EXCLUSIVE connectives the most, (2) CONT. SUBS. and HYPONYMOUS connectives have the greatest values for V, and (3) V shows the greatest sensitivity to the order of its arguments in the case of HYPONYMY.

The co-occurrence data were used to construct a Gaussian classifier, by assuming the values for D and V are generated by Gaussians.[2] First, normal functions were used to calculate the likelihood ratio of p and q being in the two relationships:

$$\frac{P(syn|data)}{P(hyp|data)} = \frac{P(syn)}{P(hyp)} \cdot \frac{P(data|syn)}{P(data|hyp)} \quad (8)$$

$$= 1 \cdot \frac{n(max(D_1, D_2); \mu_{syn}, \sigma_{syn})}{n(max(D_1, D_2); \mu_{hyp}, \sigma_{hyp})} \quad (9)$$

[2]KL divergence is right skewed, so a log-normal model was used to model D, whereas a normal model used for V.

where $n(x; \mu, \sigma)$ is the normal function with mean μ and standard deviation σ, and where μ_{syn}, for example, denotes the mean of the Gaussian model for SYNONYMY. Next the likelihood ratio for p and q was divided by that for p and q'. If this value was greater than 1, the model predicted p and q were SYNONYMS, otherwise HYPONYMS. The same technique was used for the second task.

4.3.2 Results

A leave-one-out cross validation procedure was used. For each triple $\langle p, q, q' \rangle$, the data concerning the pairs p, q and p, q' were held back, and the remaining data used to construct the models. The results are shown in Table 4. For comparison, a random baseline classifier achieves 50% accuracy.

The results demonstrate the utility of the new variance-based function V. The new variance-based function V is better than KL divergence at distinguishing HYPONYMY from SYNONYMY ($\chi^2 = 11.13, df = 1, p < 0.001$), although it performs worse on the coarser grained task. This is consistent with the expectations of Table 1. The two classifiers were also combined by making a naive Bayes assumption. This gave an accuracy of 76.1% on the first task, which is significantly better than just using KL divergence ($\chi^2 = 5.65, df = 1, p < 0.05$), and not significantly worse than using V. The combination's accuracy on the second task was 76.2%, which is about the same as using KL divergence. This shows that combining similarity- and variance-based measures can be useful can improve overall performance.

5 Conclusions

The concepts of lexical similarity and substitutability are of central importance to psychology, artificial intelligence and computational linguistics.

To our knowledge this is the first modelling study of how these concepts relate to lexical items involved in discourse-level phenomena. We found a three way correspondence between data sources of quite distinct types: distributional similarity scores obtained from lexical co-occurrence data, substitutability judgements made by linguists, and the similarity ratings of naive subjects.

The substitutability of lexical items is important for applications such as text simplification, where it can be desirable to paraphrase one discourse connective using another. Ultimately we would like to automatically predict substitutability for individual tokens. However predicting whether one connective can either a) always, b) sometimes or c) never be substituted for another is a step towards this goal. Our results demonstrate that these general substitutability relationships have empirical correlates.

We have introduced a novel variance-based function of two distributions which complements distributional similarity. We demonstrated the new function's utility in helping to predict the substitutability of connectives, and it can be expected to have wider applicability to lexical acquisition tasks. In particular, it is expected to be useful for learning relationships which cannot be characterised purely in terms of similarity, such as hyponymy. In future work we will analyse further the empirical properties of the new function, and investigate its applicability to learning relationships between other classes of lexical items such as nouns.

Acknowledgements

I would like to thank Mirella Lapata, Alex Lascarides, Alistair Knott, and the anonymous ACL reviewers for their helpful comments. This research was supported by EPSRC Grant GR/R40036/01 and a University of Sydney Travelling Scholarship.

References

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

James R. Curran and M. Moens. 2002. Improvements in automatic thesaurus extraction. In *Proceedings of the Workshop on Unsupervised Lexical Acquisition*, Philadelphia, USA.

Gregory Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, Boston.

Brigitte Grote and Manfred Stede. 1998. Discourse marker choice in sentence planning. In Eduard Hovy, editor, *Proceedings of the Ninth International Workshop on Natural Language Generation*, pages 128–137, New Brunswick, New Jersey. Association for Computational Linguistics.

M. Halliday and R. Hasan. 1976. *Cohesion in English*. Longman.

Jerry A Hobbs. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Ben Hutchinson. 2004a. Acquiring the meaning of discourse markers. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, pages 685–692.

Ben Hutchinson. 2004b. Mining the web for discourse markers. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, pages 407–410, Lisbon, Portugal.

Ben Hutchinson. 2005. Modelling the similarity of discourse connectives. To appear in *Proceedings of the the 27th Annual Meeting of the Cognitive Science Society (CogSci2005)*.

Andrew Kehler. 2002. *Coherence, Reference and the Theory of Grammar*. CSLI publications.

Alistair Knott. 1996. *A data-driven methodology for motivating a set of coherence relations*. Ph.D. thesis, University of Edinburgh.

Mirella Lapata and Alex Lascarides. 2004. Inferring sentence-internal temporal relations. In *In Proceedings of the Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics Annual Meeting*, Boston, MA.

Lillian Lee. 1999. Measures of distributional similarity. In *Proceedings of ACL 1999*.

Daniel Marcu and Abdessamad Echihabi. 2002. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, Philadelphia, PA.

Daniel Marcu. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. The MIT Press.

Scott McDonald. 2000. *Environmental determinants of lexical processing effort*. Ph.D. thesis, University of Edinburgh.

George A. Miller and William G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.

M. Moser and J. Moore. 1995. Using discourse analysis and automatic text generation to study discourse cue usage. In *Proceedings of the AAAI 1995 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*.

Philip Resnik and Mona Diab. 2000. Measuring verb similarity. In *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society*, Philadelphia, US, August.

Philip Resnik. 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130.

H. Rubenstein and J. B. Goodenough. 1965. Contextual correlates of synonymy. *Computational Linguistics*, 8:627–633.

Advaith Siddharthan. 2003. Preserving discourse structure when simplifying text. In *Proceedings of the 2003 European Natural Language Generation Workshop*.

Julie Weeds and David Weir. 2003. A general framework for distributional similarity. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, Sapporo, Japan, July.

Sholom M. Weiss and Casimir A. Kulikowski. 1991. *Computer systems that learn*. Morgan Kaufmann, San Mateo, CA.