June 2008

Easily Identifiable Discourse Relations

Emily Pitler
University of Pennsylvania, epitler@seas.upenn.edu

Mridhula Raghupathy
University of Pennsylvania, mridhula@seas.upenn.edu

Hena Mehta
University of Pennsylvania, hena@seas.upenn.edu

Ani Nenkova
Univesity of Pennsylvania, nenkova@seas.upenn.edu

Alan Lee
University of Pennsylvania, aleewk@seas.upenn.edu

See next page for additional authors

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-24.

Easily Identifiable Discourse Relations

Abstract
We present a corpus study of local discourse relations based on the Penn Discourse Tree Bank, a large manually annotated corpus of explicitly or implicitly realized contingency, comparison, temporal and expansion relations. We show that while there is a large degree of ambiguity in temporal explicit discourse connectives, overall discourse connectives are mostly unambiguous and allow high accuracy classification of discourse relations. We achieve 93.09% accuracy in classifying the explicit relations and 74.74% accuracy overall. In addition, we show that some pairs of relations occur together in text more often than expected by chance. This finding suggest that global sequence classification of the relations in text can lead to better results, especially for implicit relations.

Comments
University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-24.

Author(s)
Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkova, Alan Lee, and Aravind K. Joshi

Easily Identifiable Discourse Relations

Emily Pitler and **Mridhula Raghupathy** and **Hena Mehta** and **Ani Nenkova** and **Alan Lee** and **Aravind Joshi**
Department of Computer Science
and
Institute for Research in Cognitive Science,
University of Pennsylvania
Philadelphia, PA 19104
{epitler, mridhula, hena, nenkova, aleewk, joshi}@seas.upenn.edu

Abstract

We present a corpus study of local discourse relations based on the Penn Discourse Tree Bank, a large manually annotated corpus of explicitly or implicitly realized contingency, comparison, temporal and expansion relations. We show that while there is a large degree of ambiguity in temporal explicit discourse connectives, overall discourse connectives are mostly unambiguous and allow high-accuracy classification of discourse relations. We achieve 93.09% accuracy in classifying the explicit relations and 74.74% accuracy overall. In addition, we show that some pairs of relations occur together in text more often than expected by chance. This finding suggest that global sequence classification of the relations in text can lead to better results, especially for implicit relations.

1 Introduction

Discourse relations between textual units are considered key for the ability to properly interpret or produce discourse. Various theories of discourse have been developed (Moore and Wiemer-Hastings, 2003) and different relation taxonomies have been proposed (Hobbs, 1979; McKeown, 1985; Mann and Thompson, 1988; Knott and Sanders, 1998). Among the most cognitively salient relations are causal (contingency), contrast (comparison), and temporal. Very often, the discourse relations are *explicit*, signaled directly by the use of appropriate discourse connectives:

(E1) He is very tired **because** he played tennis all morning.

(E2) He is not very strong, **but** he can run amazingly fast.

(E3) We had some tea in the afternoon and **later** went to a restaurant for a big dinner.

Discourse relations can also be *implicit*, inferred by the context of the utterance and general world knowledge.

(I1) I took my umbrella this morning. [**because**] The forecast was rain in the afternoon.

(I2) She is never late for meetings. [**but**] He always arrives 10 minutes late.

(I3) She woke up early. [**afterward**] She had breakfast and went for a walk in the park.

An additional complication for automatic classification of discourse relations is that even in the presence of an explicit discourse connective, the connective might be ambiguous between several senses. For example, *since* can be used to signal either a temporal or a contingency relation.

They have not spoken to each other since the huge argument they had last fall.

Since you never replied to the invitation, I assumed you were not coming.

Several questions arise that are directly related to efforts in automatic recognition of discourse relations.

In a general text, what is the proportion of explicit versus implicit relations? Since implicit relations are presumably harder to recognize, the larger their proportion, the more difficult the overall discourse relation assignment in text would be.

How ambiguous are discourse connectives? The degree of ambiguity would give an upper bound on the accuracy with which explicit relations can be identified. The more ambiguous discourse connectives are, the more difficult it would be to automatically decide which discourse relation is expressed in a given sentence, even in the presence of a connective.

In a text, are adjacent discourse relations independent of each other or are certain sequences of relations more likely? In the latter case, the "discourse grammar" of text can be used and easy to identify relations such as unambiguous explicit relations can help determine the class of implicit relations that immediately follow or precede them.

In this study, we address the above questions using the largest existing corpus manually annotated with discourse relations—the Penn Discourse Tree Bank (Prasad et al., 2008). Our work complements data intensive approaches that use heuristics in order to circumvent the problem of expensive-to-obtain annotations (Marcu and Echihabi, 2002; Lapata and Lascarides, 2004; Sporleder and Lascarides, 2005; Blair-Goldensohn et al., 2007). These approaches are based on the idea that unambiguously marked explicit discourse relations can be used to learn classifiers for implicit relations. Unambiguous examples are collected from a large corpus, for example sentences containing "because" are extracted as representative causal relations and sentences containing the connective "but" are extracted to represent contrast relations. The performance of the resulting classifiers is tested without the need for manual annotations, using a clever technique—deleting the connective from the sentence and predicting the discourse relations based on features different from the discourse connective itself.

While such approaches are very flexible and do not need expensive manual annotations, recent studies have lead to the conclusion that the reported classification performance might be misleading (Sporleder and Lascarides, in press). In their study Sporleder and Lascaridies demonstrate that discourse relation classifiers trained on unambiguous explicit examples do not perform well when tested on actual, hand-annotated, *implicit* examples. Moreover, the data-intensive unsupervised approaches leave open questions such as what is the overall ambiguity of discourse connectives (only a handful of unambiguous ones are used) and what are the relative proportions of explicit and implicit relations of a given type. It is in general possible that implicit relations are more rare, and that inferring the relation without the help of any connective makes the task artificially difficult. In this papers we answer these two questions, quantifying the degree of ambiguity in discourse connectives overall (Section 3) and showing that reasonable performance of classification of discourse connectives can be achieved, based on the overt discourse connectives alone (Section 4). Finally, we show that the easy to disambiguate explicit discourse relations can be helpful in identifying implicit discourse relations since there are patterns in the sequences of relations in text (Section 5).

2 The Penn Discourse Tree Bank

The Penn Discourse Treebank (PDTB) is a new resource (Prasad et al., 2008) of annotated discourse relations along with their semantic classifications. [1] The annotation covers the same 1 million word Wall Street Journal (WSJ) corpus used for the Penn Treebank (Marcus et al., 1994), although the parse trees of the Penn Treebank were not used to constrain the discourse annotation process.

The PDTB is the first corpus with both explicit and implicit discourse relations annotated for the same texts. By definition, an explicit relation is triggered by the presence of a discourse connective which occurs overtly in the text. The discourse connective can essentially be viewed as a discourse-level predicate which takes two clausal arguments. The corpus recognizes 100 such explicit connectives and contains annotations for 19,458 explicit relations [2].

The PDTB also contains provisions for the annotation of implicit discourse relations which are inferred by the reader but are not overtly marked by a discourse connective. An *implicit relation* is assumed to be constrained by adjacency— it is only inferred between two adjacent sentences, and only in the absence of an explicit discourse connective. The same set of discourse relation types was used for both implicit and explicit relations.

[1] The PDTB also contains annotation for attribution, which contains features indicating how and to whom a discourse relation and its arguments are attributed. Unlike most other approaches to discourse annotation, attribution is not treated as a discourse relation in the PDTB. Attribution will not be further discussed in this paper.

[2] The PDTB allows annotators to tag a relation with multiple senses. In this work we count both of the annotated senses. So even though there are only 18,459 explicit relations, there are 19,458 explicit senses.

There are a total of 16,584 implicit relations annotated in the corpus. [3]

It should be emphasized here that each discourse relation is always associated with exactly two arguments. In the case of an explicit relation, one of the arguments is always syntactically associated with the explicit connective. The other argument is unconstrained as to its location in the text. For implicit relations, the two arguments must be structurally adjacent.

The PDTB also contains annotations of three other less common types in cases where an implicit relation could not be inferred.

An *AltLex* relation (which stands for Alternative Lexicalization) is annotated between two adjacent sentences when there is some structural pattern in the second sentence which is not a discourse connective but signals the presence of a discourse relation. In such cases, the insertion of an implicit connective would have lead to some redundancy in the expression of the relation.

> And she further stunned her listeners by revealing her secret garden design method: Commissioning a friend to spend five or six thousand dollars . . . on books that I ultimately cut up. **AltLex [After that],** the layout had been easy.

An *EntRel* relation (Entity Relation) is annotated when two adjacent sentences are related only because of the mention of the same discourse entity and not by a discourse relation.

A *NoRel* (No Relation) is annotated when none of the above relations can be inferred.

In this paper, we focus on explicit and implicit relations as they make up the vast majority of the corpus. In what follows, we will consider AltLex, EntRel, and NoRel to be part of an *Other* category.

In addition to discourse relations and their arguments, the PDTB also provides the *senses* of each relation(Miltsakaki et al., 2008). The tagset of senses are organized hierarchically into three levels - class, type, and subtype. The top class level contains the four major semantic classes: Expansion, Comparison, Contingency and Temporal. Briefly, Expansion covers those relations where the second argument expands the discourse of the first argument or move its narrative forward. Comparison relations highlight prominent differences between the two arguments of a relation. Contingency is marked when one of the situations described in an argument causally influences the other argument. Temporal relations are marked when the situations described in the arguments are related temporally, either synchronously or sequentially (PDTB-Group, 2008).

[3] Again, because of multiple senses per relation, the 16,584 senses are part of 16,224 relations.

Class	Explicit (%)	Implicit (%)	Total
Comparison	5590 (69.05%)	2505 (30.95%)	8095
Contingency	3741 (46.75%)	4261 (53.25%)	8002
Temporal	3696 (79.55%)	950 (20.45%)	4646
Expansion	6431 (42.04%)	8868 (57.96%)	15299

Table 1: Discourse relation distribution in semantic and explicit/implicit classes of the 34,512 discourse relations in PDTB

Each of these four major classes are further divided into types and subtypes with more refined semantic definitions. For example, the Comparison class contains two types, Contrast and Concession, and within Contrast there are two subtypes Juxtaposition and Opposition. In our experiments, we chose to use only the top class level of the semantic hierarchy, restricting ourselves therefore to the four major classes of Expansion, Comparison, Contingency, and Temporal. We assume that for most applications, the distinction between whether two sentences are related temporally or contingently will be more crucial than finer-grained distinctions such as whether the relation is juxtaposition or opposition.

Table 1 shows the distribution of discourse relations between the four main relation classes and their type of realization (implicit or explicit). Interestingly, temporal and comparison relations are predominantly explicit. About 80% and 70%, respectively, of their occurrences are marked by a discourse connective. The contingency relations are almost evenly distributed between explicit and implicit. The expansion relations, the overall largest class of discourse relations, is in most cases implicit and not marked by a discourse connective.

Given the figures in Table 1, we would expect that overall temporal and comparison relations will be more easily identified since they are overtly marked. Of course this would only be the case if discourse markers are mostly unambiguous.

3 Ambiguity of discourse connectives

Here we show all connectives that appear more than 50 times in the PDTB, their predominant sense (comparison, contingency, temporal or expansion), as well as the percentage of occurrences of the connective in its predominant sense. For example the connective *but* has *comparison* as its predominant sense and 97.19% of the 3,308 occurrences of this connective were in the comparison sense.

Comparison *but* (3308; 97.19%), *while* (781; 66.07%), *however* (485; 99.59%), *although* (328; 99.70%), *though* (320; 100.00%), *still* (190; 98.42%), *yet* (101; 97.03%)

Expansion *and* (3000; 96.83%), *also* (1746; 99.94%), *for example* (196; 100.00%), *in addition* (165; 100.00%), *instead* (112; 97.32%), *indeed* (104; 95.19%), *moreover* (101; 100.00%), *for instance* (98, 100.00%), *or* (98; 96.94%), *unless* (95; 98.95%), *in fact* (82; 92.68%) *separately* (74; 100.00%)

Contingency *if* (1223; 95.99%), *because* (858, 100.00%), *so* (263; 100.00%), *since* (184; 52.17%), *thus* (112; 100.00%), *as a result* (78; 100.00%)

Temporal *when* (989; 80.18%), *as* (743; 70.26%), *after* (577; 99.65%), *then* (340; 93.24%), *before* (326; 100.00%), *meanwhile* (193; 48.70%), *until* (162; 87.04%), *later* (91; 98.90%), *once* (84; 95.24%)

The connectives that signal comparison and contingency are mostly unambiguous. Obvious exceptions are two of the connectives that are often used to signal temporal relations: *while* and *since*. The predominant senses of these connectives are comparison (66.07%) and contingency (52.17%) respectively. Disambiguating these problematic connectives has already been addressed in previous work (Miltsakaki et al., 2005), but even the predominantly temporal connectives are rather ambiguous. For example less than 95% of the occurrances of *meanwhile, as, when, until,* and *then* are temporal relaions.

We give some examples of these ambiguities in the sentences below.

Comparison: While U.S. officials voice optimism about Japan's enlarged role in Asia, they also convey an undertone of caution.

Temporal: While giving the Comprehensive Test of Basic Skills to ninth graders at Greenville High School last March 16, she spotted a student looking at crib sheets.

Contingency: Vicar Marshall admits to mixed feelings about this issue, **since** he is both a vicar and an active bell-ringer himself.

Temporal: Since chalk first touched slate, schoolchildren have wanted to know: What's on the test?

Considering all connectives in the corpus, they appear in their predominant sense 93.43% (for comparsion), 94.72% (for contingency), 84.10% (for temporal), and 97.63% (for expansion) of the time. Temporal connectives are most ambiguous and connectives signaling expansion are least ambiguous.

We have so far concentrated on the ambiguity between different types of relations. Based on the data above, one might think that one could choose the connectives that almost always correspond to a particular sense (for example *and* is almost always an Expansion) and use these words to find the explicit relations. However, this view may be too optimistic. There is another type of ambiguity–words may be ambiguous as to whether or not they serve as a discourse connective. For example, consider the following two uses of *and*.

- Selling picked up as previous buyers bailed out of their positions *and* aggressive short sellers – anticipating further declines – moved in.
- My favorite colors are blue *and* green.

In the first sentence, "and" is being used as a discourse connective, whereas in the second sentence, "and" is simply being used to join two adjectives, and is not marking an explicit expansion discourse relation.

Prior work has reported that disambiguating between discourse and general uses based on syntactic features can be performed with high precision and recall (above 0.95). Still, of the 100 cue phrases for discourse relations in the PDTB, only 11 of them appear as a discourse connective more than 90% of the time (*although*, *in turn*, *afterward*, *consequently*, *additionally*, *alternatively*, *whereas*, *on the contrary*, *if and when*, *lest*, and *on the one hand...on the other hand*). There is quite a range among the most frequent connectives: *although* appears as a discourse connective 91.4% of the time, while *or* only serves a discourse function 2.8% of the times it appears.

Nevertheless, the percentages might be underestimated in some cases. For example, certain connectives such as *because* or *instead* have corresponding counterparts which take nominalized arguments (*because of*, *instead of*). Others like *until* or *since* might also sometimes take a nominalized complement rather than a clausal complement. For practical reasons, many of these cases are not annotated in the PDTB and annotators were instructed to look primarily for *clausal* arguments to discourse connectives (PDTB-Group, 2008). If these instances of connectives taking nominalized arguments are accounted for, the ratio of connective:non-connective for these expressions will obviously be higher.

The figures above should be kept in mind when assessing the results of classifying relation based on the explicit connective only, which we turn to in the next section.

4 Automatic classification of discourse relations

The analyses in the previous sections show two very positive trends: many of the discourse relations are explicitly marked by the use of a discourse connective, especially comparison and temporal relations, and discourse connectives are overall mostly unambiguous. These facts would suggest that even based only on the connective, classification of discourse relations could be done well for all data (including both implicit and explicit examples) and particularly well for explicit examples alone. Indeed, Tables 2 and 3 show the performance of a decision tree classifier for discourse relations, on all data and on the explicit subset respectively. Rather than down-sampling the data to obtain an even number of examples for each type of relations as has been done in prior studies (Marcu and Echihabi, 2002; Sporleder and Lascarides, 2005), we use the natural distribution of relation classes found in the Wall Street Journal texts.

The tables show the classification accuracy, precision and recall for a decision tree classifier, using 10-fold cross validation. There are four task settings, distinguishing each type of relation from all others. For example, comparison relations can be distinguished from all other relations in the corpus with overall accuracy of 91.28%, based only on the discourse connective (first entry in Table 2). The recall for recognizing comparison relations is 0.66, directly reflecting the fact that 31% of all comparison relations are implicit and the connective feature did not help at all in those cases. Over explicit data only (Table 3), the classification accuracy for comparison relation versus any other relation is 97.23%, and precision and recall is 0.95 and above.

As expected, the overall accuracy of identifying contingency and expansion relations is lower, 84.44% and 77.51% on all data respectively, reflecting the fact that these relations are often implicit. But by themselves these accuracy numbers are actually reasonable, setting a rather high baseline for any more sophisticated method tackling implicit relations. On explicit data only (table 3), the binary classification accuracy for the four main types of relations is 94% and higher with excellent precision and recall.

In four-way classification, disambiguating between the four main semantic types of discourse relations leads to 74.74% classification accuracy. The accuracy for four-way classification of explicit relations is 93.09%. The precision and recall for each class is shown in Table 4. The worst performance on the explicit portion of the data is the precision for temporal relations and the recall for contingency relations, both of which are 0.84.

Task	Class	Accuracy (majority)	Precision, Recall
Comparison		91.28% (76.54%)	
	Comp.		0.947, 0.665
	Not Comp.		0.906, 0.989
Contingency		84.44% (76.81%)	
	Cont.		0.954, 0.345
	Not Cont.		0.834, 0.995
Temporal		94.79% (86.54%)	
	Temp.		0.885, 0.705
	Not Temp.		0.955, 0.986
Expansion		77.51% (55.67%)	
	Exp.		0.666, 0.986
	Not Exp.		0.982, 0.608

Table 2: Decision tree classification accuracy, precision and recall for classification of all relations (implicit and explicit) using only the connective.

Task	Class	Accuracy (majority)	Precision, Recall
Comparison		97.23% (69.72%)	
	Comp.		0.947, 0.962
	Not Comp.		0.983, 0.977
Contingency		93.99% (79.73%)	
	Cont.		0.954, 0.739
	Not Cont.		0.937, 0.991
Temporal		95.4% (79.98%)	
	Temp.		0.885, 0.885
	Not Temp.		0.971, 0.971
Expansion		97.61% (65.16%)	
	Exp.		0.974, 0.957
	Not Exp.		0.977, 0.986

Table 3: Decision tree classification performance on *explicit* data only

Class	Precision	Recall
Temporal	0.841 [0.841]	0.729 [0.903]
Expansion	0.658 [0.973]	0.982 [0.957]
Contingency	0.948 [0.947]	0.369 [0.844]
Comparison	0.935 [0.935]	0.671 [0.971]

Table 4: Four-way classification. The first number is for all data, second for explicit relations only.

Class	E.Exp	E.Comp	E.Cont	E.Temp
I.Exp	**.17**	.15	.09	.09
I.Comp	**.15**	.12	.10	.10
I.Cont	.16	**.20**	.11	.09
I.Temp	.14	.10	.08	**.15**

Table 5: Probability of each implicit relation being preceded by each explicit relation.

Class	E.Exp	E.Comp	E.Cont	E.Temp
I.Exp	**.17**	.14	.08	.08
I.Comp	.14	**.16**	.11	.09
I.Cont	**.16**	**.16**	.13	.09
I.Temp	.13	.12	.09	**.18**

Table 6: Probability of each implicit relation being followed by each explicit relation.

5 N-gram discourse relation models

We have shown above that some relations, such as comparison, can be easily identified because they are often explicit and use an unambiguous connective. However, one must build a more subtle automatic classifier to find the implicit relations. We now look at the frequencies in which various relations are adjacent in the PDTB. Results from previous studies of discourse relations suggest that the *context* of a relation can be helpful in disambiguating the relation (Wellner et al., 2006). Here we identify specific dependencies that exist between sequences of relations in human-written text.

We computed the transitional probabilities in both directions for each pair of relations. For the forward direction, we computed:

$$P(R_i|R_{i-1}) = \frac{Count(R_{i-1}, R_i)}{Count(R_{i-1})} \quad (1)$$

For the backwards direction, we computed:

$$P(R_i|R_{i+1}) = \frac{Count(R_i, R_{i+1})}{Count(R_{i+1})} \quad (2)$$

As one can see in Tables 5 and 6, there is a distinct pattern for each type of relation, albeit a weak one. Explicit expansions are most likely to be adjacent to implicit expansions, explicit comparisons are most likely to follow implicit comparisons, explicit comparisons are most likely to be adjacent to implicit contingencies, and explicit temporals are most likely to be adjacent to implicit temporals. These results suggest that the neighboring explicit relations may be a useful feature for automatic classifiers of implicit relations.

First Relation	Second Relation	χ^2	p-value
Other	Other	66.2	$<$ .000001
Other	I. Expansion	30.5	$<$ .000001
I. Expansion	Other	20.2	.000007
E. Comparison	**I. Contingency**	**20.1**	**.000007**
E. Comparison	E. Comparison	17.4	.000030
I. Contingency	Other	14.8	.000120
Other	I. Contingency	13.6	.000228
E. Comparison	Other	13.3	.000262
E. Comparison	**I. Expansion**	**9.91**	**.001614**
I. Temporal	**E. Temporal**	**9.42**	**.002141**
I. Contingency	**E. Contingency**	**9.29**	**.002302**
I. Expansion	I. Expansion	9.09	.002569
Other	E. Expansion	6.37	.011567
I. Expansion	**E. Expansion**	**6.34**	**.011783**
I. Temporal	I. Expansion	5.52	.018784
E. Expansion	**I. Expansion**	**5.50**	**.019050**
I. Comparison	I. Expansion	5.45	.0195
I. Contingency	**E. Comparison**	**4.95**	**.0260**
E. Temporal	E. Contingency	4.24	.039571
E. Contingency	Other	4.15	.041728
I. Expansion	I. Contingency	3.93	.047475

Table 7: χ^2 results for pairs of relations

We also computed χ^2 statistics to test the independence of each pair of relations. The question is: do relations A and B occur adjacent to each other more than they would simply due to chance? The pairs of relations which have significant associations with each other ($pval < 0.05$) are shown in Table 7 [4]. Note that seven of these pairs consist of one implicit and one explicit relation (we have highlighted these pairs in bold). For example, explicit comparison and implicit contingency co-occur much more often than would be expected if they were independent. As explicit comparisons are generally fairly easy to identify, knowing that they tend to co-occur may be helpful when searching for implicit contingency relations in a text.

5.1 Perplexity

We examined the perplexities of n-gram models of discourse relations. Perplexity is often characterized as the average branching factor and is defined as $2^{entropy}$, or

$$2^{-\sum_{i=1}^{N} \frac{1}{N} \log_2 P(x_i|Model)} \quad (3)$$

If the context of a relation is helpful in disambiguating the sense of the relation, perplexities of the bigram and trigram models would be significantly lower than the perplexity of the unigram

[4] The significance of (Other, Other) should not be surprising, as it is driven by pairs of (EntRel, EntRel), and (Barzilay and Lapata, 2005) showed entity-based coherence is an important part of discourse coherence.

Model	Top-level Perplexity	All-levels Perplexity
Baseline	10	74
Unigram	8.687	24.1
Bigram	8.462	21.9
Trigram	8.305	14.0

Table 8: Perplexities of ngram discourse models over explicit/implicit discourse relations

model–this would imply that knowing the previous relations makes predicting the current relation much easier. At the top level of semantic relations, it turns out that the decrease of perplexity is indeed present, but not very large. If one is trying to choose among Implicit/Explicit Expansion/Comparison/Contingency/Temporal, Other, and End (for end of document), then if one assumes a purely uniform distribution, the average branching factor is ten (since we have ten choices). If we use the proportion of each type of relation (unigram model), we reduce our perplexity to 8.687. Knowing the label of the previous relation produces only a small improvement (perplexity = 8.462). Knowing the previous two relations also does not help very much over the bigram model (perplexity of 8.305).

However, in the above discussion we were looking only at the top level sense classes. The connectives in the PDTB are actually annotated according to a three-tiered hierarchy of senses. For example, the *Expansion* relation is subdivided into *Conjunction*, *Instantiation*, *Restatement*, *Alternative*, *Exception*, and *List*. Perhaps it is the case that there are dependencies between relations, but they are more obvious at some lower level. In order to test this, we computed the perplexities using the finer-grained classification of discourse relations.

The results are shown in Table 8. In the finer-grained case, adding more context does reduce the perplexity considerably. The average branching factor for predicting the full annotation is 24.1 given only the unigram model, but decreases to 14.0 given the previous two labels. Thus, even if one is mostly concerned with the top-level sense class, it may be useful to predict the full annotation, as it is more helpful for predicting the surrounding relations.

Given our finding that fine-grained relations provide better context, we then repeated the χ^2 test for significant dependencies between pairs of fine-grained relations. The results are shown in Table 9. Again, we have shown in bold the explicit/implicit combinations that exhibit significant associations. Whereas before we had seen that implicit contingencies were most likely after explicit comparisons, we now know that they are especially likely after explicit contrasts and explicit contra-expectation concessions. Thus, being able to find those fine-grained relations may help in finding implicit contingency relations. We can see also that prior to explicit expansions, we are more likely to have implicit expansions of the same type: the implicit list relation is more likely to occur immediately prior to an explicit list, and implicit conjunction is more likely prior to an explicit conjunction.

6 Conclusion

Overall, we have tried to summarize the difficulty of finding discourse relations using the Penn Discourse Treebank. We noted that explicit and implicit relations are approximately evenly distributed overall, making the task easier than many researchers have feared. We have found that some relations, such as temporal and comparison, are more likely to be explicit than implicit, making them relatively easier to find, while the contingency (causal) relation is more often implicit. Among the discourse connectives, the majority are not very ambiguous between the different types of relations, with some notable exceptions such as *since* and *meanwhile*.

We have also analyzed the ambiguity of specific cue words–given that a word such as *because* appears, how likely is it to be acting as a discourse connective? We found that the ambiguity varies widely with the cue word. While *additionally* was used as a connective 100% of the time it appeared in the WSJ Corpus, at the other end of the spectrum, *for* was only used as a connective .03% of the time. The discourse versus non-discourse usage can be largely disambiguated on syntactic grounds, but we leave for future work a more detailed study of predicting whether a connective is being used in a discourse sense.

We have carried out a novel quantitative study of the patterns of dependencies between discourse relations. We found that while there does not appear to be a clear template for the sequence of relations, there are individual relation pairs that tend to co-occur. Specifically, we found that even

First Relation	Second Relation	χ^2	p-value
Other	Other	66.2	$< .000001$
I. Expansion.List	I. Expansion.List	34.3	$< .000001$
I. Contingency.Cause.Reason	Other	9.19	0.002434
I. Expansion.List	**E. Expansion.List**	**9.14**	**0.002501**
I. Expansion.Instantiation	Other	6.94	0.008429
E. Comparison.Contrast	**I. Contingency.Cause.Reason**	**6.34**	**0.011805**
I. Expansion.Instantiation	I. Expansion.Conjunction	6.31	0.012006
I. Expansion.Conjunction	I. Expansion.Conjunction	6.24	0.012490
Other	I. Contingency.Cause.Reason	4.57	0.032536
I. Expansion.Conjunction	**E. Expansion.Conjunction**	**4.31**	**0.037889**
E. Comparison.Concession.Contra-expectation	**I. Contingency.Cause.Reason**	**3.92**	**0.047715**

Table 9: χ^2 results for pairs of fine-grained relations

though contingency relations are likely to be implicit and thus difficult to find, they are likely to be found near an explicit comparison. We plan to exploit these findings in future work, addressing discourse relation labeling in text as a sequence labeling problem and using the explicit cue words of surrounding relations as features for finding the "hidden" implicit relations.

References

Barzilay, R. and M. Lapata. 2005. Modeling local coherence: an entity-based approach. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 141–148.

Blair-Goldensohn, S., K. McKeown, and O. Rambow. 2007. Building and refining rhetorical-semantic relation models. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 428–435.

Hobbs, J. 1979. Coherence and coreference. *Cognitive Science*, 3:67–90.

Knott, A. and T. Sanders. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30(2):135–175.

Lapata, M. and A. Lascarides. 2004. Inferring sentence-internal temporal relations.

Mann, W. and S. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8:243–281.

Marcu, D. and A. Echihabi. 2002. An unsupervised approach to recognizing discourse relations.

Marcus, M.P., B. Santorini, and M.A. Marcinkiewicz. 1994. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

McKeown, Kathleen R. 1985. *Text generation: using discourse strategies and focus constraints to generate natural language text*. Cambridge University Press, New York, NY, USA.

Miltsakaki, E., N. Dinesh, R. Prasad, A. Joshi, and B. Webber. 2005. Experiments on sense annotations and sense disambiguation of discourse connectives. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT2005)*.

Miltsakaki, Eleni, Livio Robaldo, Alan Lee, and Aravind Joshi. 2008. Sense annotation in the penn discourse treebank. *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science*, 4919:275–286.

Moore, J. and P. Wiemer-Hastings. 2003. Discourse in computational linguistics and artificial intelligence.

PDTB-Group. 2008. The penn discourse treebank 2.0 annotation manual. Technical Report IRCS-08-01, Insitute for Research in Cognitive Science, University of Pennsylvania.

Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*.

Sporleder, C. and A. Lascarides. 2005. Exploiting linguistic cues to classify rhetorical relations. In *Proceedings of Recent Advances in Natural Language Processing (RANLP-05)*, pages 532–539.

Sporleder, C. and A. Lascarides. in press. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*.

Wellner, B., J. Pustejovsky, C. Havasi, R. Sauri, and A. Rumshisky. 2006. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGDIAL Workshop on Discourse and Dialogue*.