

# The Effects of Discourse Connectives Prediction on Implicit Discourse Relation Recognition

Zhi Min Zhou<sup>†</sup>, Man Lan<sup>†,§</sup>, Zheng Yu Niu<sup>‡</sup>, Yu Xu<sup>†</sup>, Jian Su<sup>§</sup>

<sup>†</sup>East China Normal University, Shanghai, PRC.

<sup>‡</sup>Baidu.com Inc., Beijing, PRC.

<sup>§</sup>Institute for Infocomm Research, Singapore.

51091201052@ecnu.cn, lanman.sg@gmail.com

## Abstract

Implicit discourse relation recognition is difficult due to the absence of explicit discourse connectives between arbitrary spans of text. In this paper, we use language models to predict the discourse connectives between the arguments pair. We present two methods to apply the predicted connectives to implicit discourse relation recognition. One is to use the sense frequency of the specific connectives in a supervised framework. The other is to directly use the presence of the predicted connectives in an unsupervised way. Results on PDTB2 show that using language model to predict the connectives can achieve comparable F-scores to the previous state-of-art method. Our method is quite promising in that not only it has a very small number of features but also once a language model based on other resources is trained it can be more adaptive to other languages and domains.

## 1 Introduction

Discourse relation analysis involves identifying the discourse relations (e.g., the comparison relation) between arbitrary spans of text, where the discourse connectives (e.g., “however”, “because”) may or may not explicitly exist in the text. This analysis is one important application both as an end in itself and as an intermediate step in various downstream NLP applications, such as text summarization, question answering etc.

As discussed in (Pitler and Nenkova., 2009b), although explicit discourse connectives may have two types of ambiguity, i.e., one is discourse or non-discourse usage (“once” can be either a temporal connective or a word meaning “formerly”), the other is discourse relation sense ambiguity

(“since” can serve as either a temporal or causal connective), their study shows that for explicit discourse relations in Penn Discourse Treebank (PDTB) corpus, the most general 4 senses, i.e., Comparison (Comp.), Contingency (Cont.), Temporal (Temp.) and Expansion (Exp.), can be easily addressed by the presence of discourse connectives and a simple method only considering the sense frequency of connectives can achieve more than 93% accuracy. This indicates the importance of connectives for discourse relation recognition.

However, with implicit discourse relation recognition, there is no connective between the textual arguments, which results in a very difficult task. In recent years, a multitude of efforts have been employed to solve this task. One approach is to exploit various linguistically informed features extracted from human-annotated corpora in a supervised framework (Pitler et al., 2009a) and (Lin et al., 2009). Another approach is to perform recognition without human-annotated corpora by creating synthetic examples of implicit relations in an unsupervised way (Marcu and Echiabi, 2002).

Moreover, our initial study on PDTB implicit relation data shows that the averaged F-score for the most general 4 senses can reach 91.8% when we obtain the sense of test examples by mapping each implicit connective to its most frequent sense (i.e., sense recognition using gold-truth implicit connectives). This high F-score performance again proves that the connectives are very crucial source for implicit relation recognition.

In this paper, we present a new method to address the problem of recognizing implicit discourse relation. This method is inspired by the above observations, especially the two gold-truth results, which reveals that discourse connectives are very important signals for discourse relation recognition. Our basic idea is to recover the implicit connectives (not present in real text) between two spans of text with the use of a language

model trained on large amount of raw data without any human-annotation. Then we use these predicted connectives to generate feature vectors in two ways for implicit discourse relation recognition. One is to use the sense frequency of the specific connectives in a supervised framework. The other is to directly use the presence of the predicted connectives in an unsupervised way.

We performed evaluation on explicit and implicit relation data sets in the PDTB 2 corpus. Experimental results showed that the two methods achieved comparable F-scores to the state-of-art methods. It indicates that the method using language model to predict connectives is very useful in solving this task.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 describes our methods for implicit discourse relation recognition. Section 4 presents experiments and results. Section 5 offers some conclusions.

## 2 Related Work

Existing works on automatic recognition of implicit discourse relations fall into two categories according to whether the method is supervised or unsupervised.

Some works perform relation recognition with supervised methods on human-annotated corpora, for example, the RST Bank (Carlson et al., 2001) used by (Soricut and Marcu, 2003), adhoc annotations used by (Girju, 2003) and (Baldrige and Lascarides, 2005), and the GraphBank (Wolf et al., 2005) used by (Wellner et al., 2006).

Recently the release of the Penn Discourse TreeBank (PDTB) (Prasad et al., 2006) has significantly expanded the discourse-annotated corpora available to researchers, using a comprehensive scheme for both implicit and explicit relations. (Pitler et al., 2009a) performed implicit relation classification on the second version of the PDTB. They used several linguistically informed features, such as word polarity, verb classes, and word pairs, showing performance increases over a random classification baseline. (Lin et al., 2009) presented an implicit discourse relation classifier in PDTB with the use of contextual relations, constituent Parse Features, dependency parse features and cross-argument word pairs. Although both of two methods achieved the state of the art performance for automatical recognition of implicit discourse relations, due to lack of human-annotated

corpora, their approaches are not very useful in the real world.

Another line of research is to use the unsupervised methods on unhuman-annotated corpus.

(Marcu and Echiabi, 2002) used several patterns to extract instances of discourse relations such as contrast and elaboration from unlabeled corpora. Then they used word-pairs between arguments as features for building classification models and tested their model on artificial data for implicit relations.

Subsequently other studies attempt to extend the work of (Marcu and Echiabi, 2002). (Sporleder and Lascarides, 2008) discovered that Marcu and Echiabi's models do not perform as well on implicit relations as one might expect from the test accuracy on synthetic data. (Goldensohn, 2007) extended the work of (Marcu and Echiabi, 2002) by refining the training and classification process using parameter optimization, topic segmentation and syntactic parsing. (Saito et al., 2006) followed the method of (Marcu and Echiabi, 2002) and conducted experiments with a combination of cross-argument word pairs and phrasal patterns as features to recognize implicit relations between adjacent sentences in a Japanese corpus.

Previous work showed that with the use of some patterns, structures, or the pairs of words, relation classification can be performed using unsupervised methods.

In contrast to existing work, we investigated a new knowledge source, i.e., implicit connectives predicted using a language model, for implicit relation recognition. Moreover, this method can be applied in both supervised and unsupervised ways by generating features on labeled and unlabeled training data and then performing implicit discourse connectives recognition.

## 3 Methodology

### 3.1 Predicting implicit connectives via a language model

Previous work (Pitler and Nenkova., 2009b) showed that with the presence of discourse connectives, explicit discourse relations in PDTB can be easily identified with more than 90% F-score. Our initial study on PDTB human-annotated implicit relation data shows that the averaged F-score for the most general 4 senses can reach 91.8% when we simply map each implicit connective to

its most frequent sense. These high F-scores indicate that the connectives are very crucial source of information for both explicit and implicit relation recognition. However, for implicit relations, there are no explicitly discourse connectives in real text. This built-in absence makes the implicit relation recognition task quite difficult. In this work we overcome this difficulty by inserting connectives into the two arguments with the use of a language model.

Following the annotation scheme of PDTB, we assume that each implicit connective takes two arguments, denoted as *Arg1* and *Arg2*. Typically, there are two possible positions for most of implicit connectives, i.e., the position before *Arg1* and the position between *Arg1* and *Arg2*. Given a set of implicit connectives  $\{c_i\}$ , we generate two synthetic sentences,  $c_i+Arg1+Arg2$  and  $Arg1+c_i+Arg2$  for each  $c_i$ , denoted as  $S_{c_i,1}$  and  $S_{c_i,2}$ . Then we calculate the *perplexity* (an intrinsic score) of these sentences with the use of a language model, denoted as  $Ppl(S_{c_i,j})$ . According to the value of  $Ppl(S_{c_i,j})$  (the lower the better), we can rank these sentences and select the connectives in top  $N$  sentences as implicit connectives for this argument pair. Here the language model may be trained on any large amount of unannotated corpora that can be cheaply acquired. Typically, a large corpora with the same domain as the test data will be used for training language model. Therefore, we chose news corpora, such as North American News Corpora.

After that, we use the top  $N$  predicted connectives to generate different feature vectors and perform the classification in two ways. One is to use the sense frequency of predicted connectives in a supervised framework. The other is to directly use the presence of the predicted connectives in an unsupervised way. The two approaches are described as follows.

### 3.2 Using sense frequency of predicted discourse connectives as features

After the above procedure, we get a sorted set of predicted discourse connectives. Due to the presence of an implicit connective, the implicit discourse relation recognition task can be addressed with the methods for explicit relation recognition, e.g., sense classification based only on connectives (Pitler et al., 2009a). Inspired by their work, the first approach is to use sense frequency of pre-

dicted discourse connectives as features. We take the connective with the lowest perplexity value (i.e., top 1 connective) as the real connective for the arguments pair. Then we count the sense frequency of this connective on the training set. Figure 1 illustrates the procedure of generating predicted discourse connective from a language model and calculating its sense frequency from training data. Here the calculation of sense frequency of connective is based on the annotated training data which has labeled discourse relations, thus this method is a supervised one.

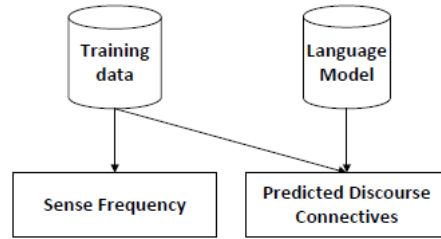


Figure 1: Procedure of generating a predicted discourse connective and its sense frequency from the training set and a language model.

Then we can directly use the sense frequency to generate a 4-feature vector to perform the classification. For example, the sense frequency of the connective *but* in the most general 4 senses can be counted from training set as 691, 6, 49, 2, respectively. For a given pair of arguments, if *but* is predicted as the top 1 connective based on a language model, a 4-dimension feature vector (691, 6, 49, 2) is generated for this pair and used for training and test procedure. Figure 2 and 3 show the training and test procedure for this method.

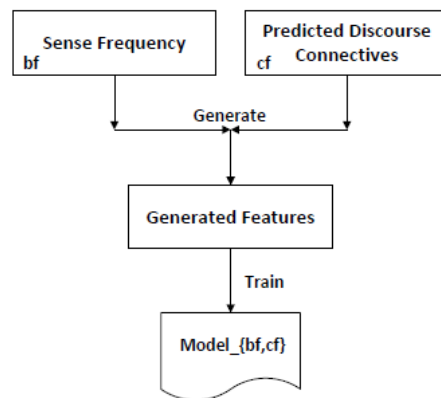


Figure 2: Training procedure for the first approach.

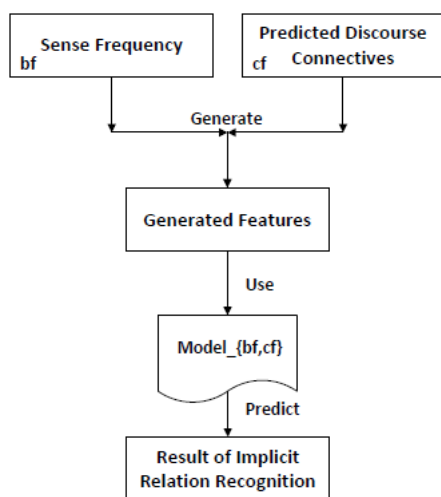


Figure 3: Test procedure for the first approach.

### 3.3 Using presence or absence of predicted discourse connective as features

(Pitler et al., 2008) showed that most connectives are unambiguous and it is possible to obtain high-accuracy in prediction of discourse senses due to the simple mapping relation between connectives and senses. Given two examples:

(E1) She paid less on her dress, *but* it is very nice.

(E2) We have to hurry up *because* the raining is getting heavier and heavier.

The two connectives, i.e., *but* in E1 and *because* in E2, convey the Comparison and Contingency senses respectively. In most cases, we can easily recognize the relation sense by the appearance of a discourse connective since it can be interpreted in only one way. That means the ambiguity of the mapping between sense and connective is quite low. Therefore, the second approach is to use only the presence of the top  $N$  predicted discourse connectives to generate a feature vector for a given pair of arguments.

## 4 Experiment

### 4.1 Data sets

We used PDTB as our data set to perform the evaluation of our methods. The corpus contains annotations of explicit and implicit discourse relations. The first evaluation is performed on the annotated implicit data set. Following the work of (Pitler et al., 2009a), we used sections 2-20 as the training set, sections 21-22 as the test set and sections 0-1 as the development set for parameter optimization (e.g.,  $N$  value). The second evaluation is per-

formed on the annotated explicit data set. We follow the method used in (Sporleder and Lascarides, 2008) to remove the discourse connective from the explicit instances and consider these processed instances as implicit ones.

We constructed four binary classifiers to recognize each main senses (i.e., Cont., Cont., Exp., Temp.) from the rest. For each sense we used equal numbers of positive and negative instances in training set. The negative instances were chosen at random from the rest of training set. For both evaluations all instances in sections 21-22 were used as test set. Table 1 lists the numbers of positive and negative instances for each sense in training, development and test sets of implicit and explicit relation data sets.

### 4.2 Evaluation and classifier

To evaluate the performance of above systems, we used two widely-used measures, F-score (i.e.,  $F_1$ ) and accuracy. In addition, in this work we used the LIBSVM toolkit to construct four linear SVM classifiers for each sense.

### 4.3 Preprocessing

We used the SRILM toolkit to build a language model and calculated the *perplexity* value for each training and test sample. The steps are described as follows. First, since perplexity is an intrinsic score to measure the similarity between training and test samples, in order to fit the restriction of perplexity we chose 3 widely-used corpora in the Newswire domain to train the language model, i.e., (1) the New York part of BLLIP North American News Text (Complete), (2) the Xin and (3) the Ltw parts of the English Gigaword Fourth Edition. For the BLLIP corpus with 1,796,386 automatically parsed English sentences, we converted the parsed sentences into original textual data. Some punctuation marks such as commas, periods, minuses, right/left parentheses are converted into their original form. For the Xin and Ltw parts, we only used the Sentence Detector toolkit in OpenNLP to split each sentence. Finally we constructed 3-, 4- and 5-grams language models from these three corpora. Table 2 lists statistics of different n-grams in the different language models and different corpora.

Next, for each instance we combined its *Arg1* and *Arg2* with connectives obtained from PDTB. There are two types of connectives, single connectives (e.g. “*because*” and “*but*”) and paral-

Table 1: Statistics of positive and negative instances for each sense in training, development and test sets of implicit and explicit relation data sets.

	Implicit				Explicit			
	Comp.	Cont.	Exp.	Temp.	Comp.	Cont.	Exp.	Temp.
Train(Pos/Neg)	1927/1927	3375/3375	6052/6052	730/730	4080/4080	2732/2732	4609/4609	2663/2663
Dev(Pos/Neg)	191/997	292/896	651/537	54/1134	438/1071	295/1214	514/995	262/1247
Test(Pos/Neg)	146/912	276/782	556/502	67/991	388/1025	235/1178	501/912	289/1124

Table 2: Statistics of different n-grams in the different language models and different corpora.

n-gram	BLLIP - New York	Gigaword-Xin	Gigaword-Ltw
1-gram	1638156	2068538	2276491
2-grams	26156851	23961796	33504873
3-grams	80876435	77799100	101855639
4-grams	127142452	134410879	159791916
5-grams	146454530	168166195	183794771

lel connectives (such as “*not only . . . , but also*”). Since discourse connectives may appear not only ahead of the *Arg1*, but also between *Arg1* and *Arg2*, we considered this case. Given a set of possible implicit connectives  $\{c_i\}$ , for a single connective  $c_i$ , we constructed two synthetic sentences,  $c_i + Arg1 + Arg2$  and  $Arg1 + c_i + Arg2$ . In case of parallel connectives, we constructed one synthetic sentence like  $c_{i,1} + Arg1 + c_{i,2} + Arg2$ .

As a result, we obtain 198 synthetic sentences ( $|c_i| * 2$  for single connective or  $|c_i|$  for parallel connective) for each pair of arguments. Then we converted all words to lower cases and used the language model trained in the above step to calculate its perplexity (the lower the better) value on sentence level. The sentences were ranked from low to high according to their perplexity scores. For example, given a sentence with arguments pair as follows:

*Arg1*: it increased its loan-loss reserves by \$93 million after reviewing its loan portfolio,

*Arg2*: before the loan-loss addition it had operating profit of \$10 million for the quarter.

we got the perplexity (*Ppl*) values for this arguments pair in combination with two connectives (*but* and *by comparison*) in two positions as follows:

1. *but* + *Arg1* + *Arg2*: *Ppl*= 349.622
2. *Arg1* + *but* + *Arg2*: *Ppl*= 399.339
3. *by comparison* + *Arg1* + *Arg2*: *Ppl*= 472.206
4. *Arg1* + *by comparison* + *Arg2*: *Ppl*= 543.051

In our second approach described in Section 3.3, we considered the combination of connectives and their position as final features like *mid\_but*, *first\_but*, where the features are binary, that is, the presence or absence of the specific connective. According to the value of  $Ppl(S_{c_i,j})$ , we tried various  $N$  values on development set to get the optimal  $N$  value.

#### 4.4 Results

Table 3 summarizes the best performance achieved using gold-truth implicit connectives, the previous state-of-art performance achieved by (Pitler et al., 2009a) and our approaches. The first line shows the result by mapping the gold-truth implicit connectives directly to the relation’s sense. The second line presents the best result of (Pitler et al., 2009a). One thing worth mentioning here is that for the Expansion relation, (Pitler et al., 2009a) expanded both training and test sets by including EntRel relation as positive examples, which makes it impossible to perform direct comparison. The third and fourth lines show the best results using our first approach, where the sense frequency is counted on explicit and implicit training set respectively. The last line shows the best result of our second approach only considering the presence of top  $N$  connectives.

Table 4 summarizes the best performance using gold-truth explicit connectives reported in (Pitler and Nenkova., 2009b) and our two approaches.

Figure 4 shows the curves of averaged F-scores on implicit connective classification with different n-gram language models. From this figure we can see that all 4-grams language models achieved around 0.5% better averaged F-score than 3-grams models. And except for Ltw corpus, other 5-grams models achieved lower averaged F-score than 4-grams models. Specially the 5-grams result of New York corpus is much lower than its 3-grams result.

Figure 5 shows the averaged F-scores of different top  $N$  on the New York corpus with 3-, 4- and 5-grams language models. The essential

Table 3: Best result of implicit relations compared with state-of-art methods.

System	Comp. vs. Not $F_1$ (Acc)	Cont. vs. Other $F_1$ (Acc)	Exp. vs. Other $F_1$ (Acc)	Temp. vs. Other $F_1$ (Acc)	Averaged $F_1$ (Acc)
Sense recognition using gold-truth implicit connectives	94.08(98.30)	98.19(99.05)	97.79(97.64)	77.04(97.07)	91.78(98.02)
Best result in (Pitler et al., 2009a)	21.96(56.59)	47.13(67.30)	76.42(63.62)	16.76(63.49)	40.57(62.75)
Use sense frequency in explicit training set	26.02(52.17)	35.72(51.70)	64.94(53.97)	13.76(41.97)	35.10(49.95)
Use sense frequency in implicit training set	24.55(63.99)	16.26(70.79)	60.70(53.50)	14.75(70.51)	29.07(64.70)
Use presence of top $N$ connectives only	21.91(52.84)	39.53(50.85)	68.84(52.93)	11.91(6.33)	35.55(40.74)

Table 4: Best result of explicit relation conversion to implicit relation compared with results using the same method.

System	Comp. vs. Not $F_1$ (Acc)	Cont. vs. Other $F_1$ (Acc)	Exp. vs. Other $F_1$ (Acc)	Temp. vs. Other $F_1$ (Acc)	Average $F_1$ (Acc)
Sense recognition using gold-truth explicit connectives in (Pitler et al., 2009a)	N/A	N/A	N/A	N/A	N/A(93.67)
Use sense frequency in explicit training set	41.62(50.96)	27.46(59.24)	48.44(50.88)	35.14(54.28)	38.17(53.84)
Use presence of top $N$ connectives only	42.92(55.77)	31.83(56.05)	47.26(55.77)	37.89(58.24)	39.98(56.46)

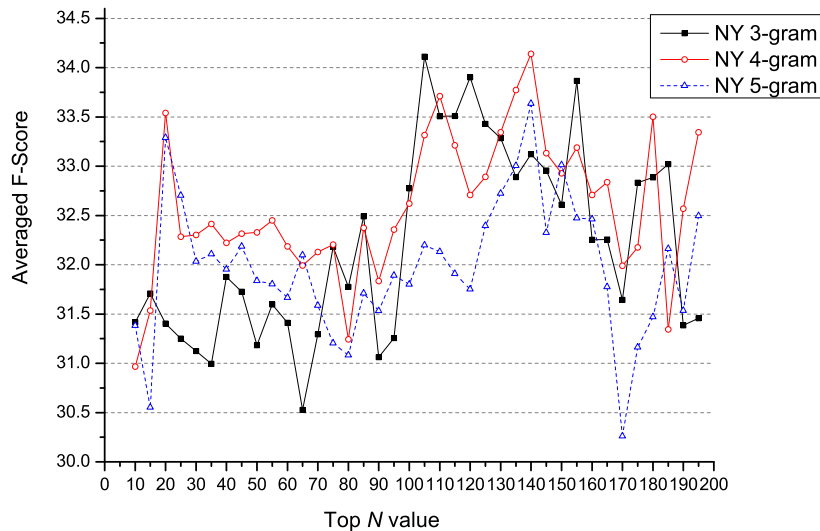


Figure 5: Curves of averages F-score on New York 3-, 4- and 5-grams language models with different top  $N$  values.

trend of these curves cannot be summarized in one sentence. But we can see that the best averaged F-scores mostly appeared in the range from 100 – 160. For 4-grams and 5-grams models, the system achieved the top averaged F-scores when  $N = 20$  as well.

#### 4.5 Discussion

Experimental results on PDTB showed that using predicted connectives achieved the comparable F-scores of the state-of-art method.

From Table 3 we can find that our results are closely to the best performance of previous state-of-art methods in terms of averaged F-score. On

the Comparison sense, our first approach has an improvement of more than 4% F-score on the previous state-of-art method (Pitler et al., 2009a). As we mentioned before, for the Expansion sense, they included EntRel relation to expand the training set and test set, which makes it impossible to perform a direct comparison. Since the positive instances size has been increased by 50%, they may achieve a higher F-score than our approach. For other relations, our best performance is slightly lower than theirs. While bearing in mind that our approach only uses a very small amount of features for implicit relation recognition. Compared

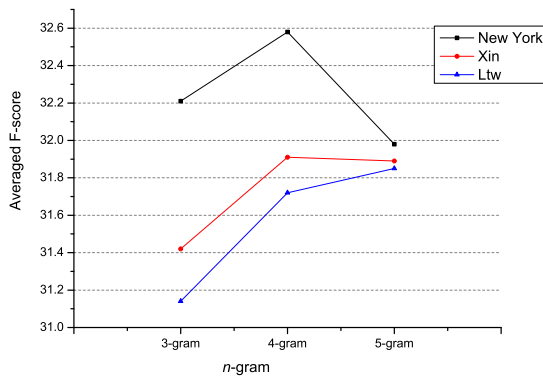


Figure 4: Curves of averaged F-score on implicit connective classification with  $n$ -Gram language model.

with other approaches involving thousands of features, our method is quite promising.

From Table 4 we observe comparable averaged F-score (39.98% F-score) on explicit relation data set to that on implicit relation data set. Previously, (Sporleder and Lascarides, 2008) also used the same conversion method to perform implicit relation recognition on different corpora and their best result is around 33.69% F-score. Although the two results cannot be compared directly due to different data sets, the magnitude of performance quantities is comparable and reliable.

By comparing with the above different systems, we find several useful observations. First, our method using predicted implicit connectives via a language model can help the task of implicit discourse relation recognition. The results are comparable to the previous state-of-art studies. Second, our method has a lot of advantages, i.e., a very small amount of features (several or no more than 200 vs. ten thousand), easy computation (only based on the trained language model vs. using a lot of NLP tools to extract a large amount of linguistically informed features) and fast running, which makes it more practical in real world application. Furthermore, since the language model can be trained on many corpora whether annotated or unannotated, this method is more adaptive to other languages and domains.

## 5 Conclusions

In this paper we have presented an approach to implicit discourse relation recognition using predicted implicit connectives via a language model.

The predicted connectives have been used for implicit relation recognition in two ways, i.e., supervised and unsupervised framework. Results on the Penn Discourse Treebank 2.0 show that the predicted discourse connectives can help implicit relation recognition and the two algorithms achieve comparable F-scores with the state-of-art method. In addition, this method is quite promising due to its simple, easy to retrieve, fast run and increased adaptivity to other languages and domains.

## Acknowledgments

We thank the reviewers for their helpful comments and Jonathan Ginzburg for his mentoring. This work is supported by grants from National Natural Science Foundation of China (No.60903093), Shanghai Pujiang Talent Program (No.09PJ1404500) and Doctoral Fund of Ministry of Education of China (No.20090076120029).

## References

- J. Baldridge and A. Lascarides. 2005. *Probabilistic head-driven parsing for discourse structure*. Proceedings of the Ninth Conference on Computational Natural Language Learning.
- L. Carlson, D. Marcu, and Ma. E. Okurowski. 2001. *Building a discourse-tagged corpus in the framework of rhetorical structure theory*. Proceedings of the Second SIG dial Workshop on Discourse and Dialogue.
- B. Dorr. LCS Verb Database. *Technical Report Online Software Database, University of Maryland, College Park, MD, 2001*.
- R. Girju. 2003. *Automatic detection of causal relations for question answering*. In ACL 2003 Workshops.
- S. Blair-Goldensohn. 2007. *Long-Answer Question Answering and Rhetorical-Semantic Relations*. Ph.D. thesis, Columbia University.
- M. Lapata and A. Lascarides. 2004. *Inferring Sentence-internal Temporal Relations*. Proceedings of the North American Chapter of the Association of Computational Linguistics.
- Z.H. Lin, M.Y. Kan and H.T. Ng. 2009. *Recognizing Implicit Discourse Relations in the Penn Discourse Treebank*. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing.
- D. Marcu and A. Echiabi. 2002. *An Unsupervised Approach to Recognizing Discourse Relations*. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.

- E. Pitler, M. Raghupathy, H. Mehta, A. Nenkova, A. Lee, A. Joshi. 2008. *Easily Identifiable Discourse Relations*. Coling 2008: Companion volume: Posters.
- E. Pitler, A. Louis, A. Nenkova. 2009. *Automatic sense prediction for implicit discourse relations in text*. Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics.
- E. Pitler and A. Nenkova. 2009. *Using Syntax to Disambiguate Explicit Discourse Connectives in Text*. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers.
- M. Porter. An algorithm for suffix stripping. In *Program*, vol. 14, no. 3, pp.130-137, 1980.
- R. Prasad, N. Dinesh, A. Lee, A. Joshi, B. Webber. 2006. *Annotating attribution in the Penn Discourse TreeBank*. Proceedings of the COLING/ACL Workshop on Sentiment and Subjectivity in Text.
- R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, B. Webber. 2008. *The Penn Discourse TreeBank 2.0*. Proceedings of LREC'08.
- M. Saito, K. Yamamoto, S. Sekine. 2006. *Using Phrasal Patterns to Identify Discourse Relations*. Proceeding of the HLTCNA Chapter of the ACL.
- R. Soricut and D. Marcu. 2003. *Sentence Level Discourse Parsing using Syntactic and Lexical Information*. Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference.
- C. Sporleder and A. Lascarides. 2008. *Using automatically labelled examples to classify rhetorical relations: an assessment*. Natural Language Engineering, Volume 14, Issue 03.
- B. Wellner, J. Pustejovsky, C. H. R. S., A. Rumshisky. 2006. *Classification of discourse coherence relations: An exploratory study using multiple knowledge sources*. Proceedings of the 7th SIGDIAL Workshop on Discourse and Dialogue.
- F. Wolf, E. Gibson, A. Fisher, M. Knight. 2005. *The Discourse GraphBank: A database of texts annotated with coherence relations*. Linguistic Data Consortium.