

# Interpretation of Chinese Discourse Connectives for Explicit Discourse Relation Recognition

Hen-Hsen Huang, Tai-Wei Chang, Huan-Yuan Chen, and Hsin-Hsi Chen

Department of Computer Science and Information Engineering

National Taiwan University

No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan

{hhhuang, twchang}@nlg.csie.ntu.edu.tw;

{b00902057, hhchen}@ntu.edu.tw

## Abstract

This paper addresses the specific features of Chinese discourse connectives, including types (word-pair and single-word), linking directions (forward and backward linking), positions and ambiguous degrees, and discusses how they affect the discourse relation recognition. A semi-supervised learning method is proposed to learn the probability distributions of discourse functions of connectives from a small labeled dataset and a big unlabeled dataset. The statistics learned from the dataset demonstrates some interesting linguistic phenomena such as connective synonyms sharing similar distributions, multiple discourse functions of connectives, and couple-linking elements providing strong clues for discourse relation resolution.

## 1 Introduction

Discourse relation labeling determines how two discourse units cohere to each other. A discourse unit may be a clause, a sentence, or a group of sentences. The labeled relation has many potential applications. Coherence is considered as a metric to evaluate the essay writing by essay scorer (Lin et al., 2011). Discourse relations are used to order sentences in an event in a summarization system (Derczynski and Gaizauskas, 2013). Sentiment transition of two clausal arguments is identified based on their discourse relation in sentiment analysis (Hutchinson, 2004; Zhou et al., 2011; Wang et al., 2012; Huang et al., 2013).

The pioneer research of discourse has been established by Hobbs (1985), Polanyi (1988), Hovy and Maier (1992), and Asher and Lascarides (1995). Various discourse relation types have been defined in the frameworks such as Sanders et al. (1992), Hovy and Maier (1992), RST-DT (Carlson et al., 2002), Wolf and Gibson (2005), and PDTB (Prasad et al., 2008). *Temporal*, *Contingency*, *Comparison*, and *Expansion*, the four classes on the top level of PDTB sense hierarchy, are common used in the discourse relation labeling tasks. When two arguments are temporally related, they form a *Temporal* relation. The *Contingency* relation talks about the situation that the event in one argument casually affects the event in the other argument. *Comparison* is used to show the difference between two arguments. The last one relation, *Expansion*, is the most common. An *Expansion* relation either expands the information for one argument in the other one or continues the narrative flow.

In the recent years, discourse relation recognition has been studied for different languages (Afanenos et al., 2012, Cartoni et al., 2013). In explicit English discourse relation labeling tasks, the accuracy of the approach using just the connectives is already quite high, 93.67%, and incorporating the syntactic features raises performance to 94.15% (Pitler and Nenkova, 2009). In our previous work, we investigate Chinese intra-sentential relation detection and show an accuracy of 81.63% and an F-score of 71.11% in the two-way classification (*Contingency* vs. *Comparison* relations) when connectives are

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

introduced as features (Huang and Chen, 2012a). We also report an accuracy of 27.10% and an F-score of 24.27% in the four-way inter-sentential relation classification when only connectives are used (Huang and Chen, 2011). Sporleder and Lascarides (2008) point out some English connectives are often ambiguous between multiple discourse relations or between discourse and non-discourse usage, and Roze et al. (2010) report the ambiguity of French connectives. This issue also occurs in Chinese. Zhou et al. (2012) propose a framework to identify the ambiguous Chinese discourse connectives, and report an F-score of 74.81% in the four-way classification at the intra-sentential level.

The above discourse relation labeling tasks are done on the datasets of different size for different languages at the intra-/inter-sentential levels, thus the results cannot be compared directly. However, these works show a tendency: discourse connectives are useful clues for explicit discourse relation recognition, and the uses of Chinese connectives in discourse relation labeling are more challenging than those of English connectives. In comparison with English, the connectives in Chinese are more and their parts of speech are diverse. There are 100 English explicit connectives annotated in the PDTB 2.0. In Chinese, the linguists report a list of 808 discourse connectives (Cheng and Tian, 1989; Cheng, 2006). In addition, the Chinese discourse connectives have a variety of parts-of-speech. For example, 假設 (jiǎ shè, suppose) is a verb and listed as a discourse connective of the *Contingency* relation.

The following examples address some specific features of Chinese discourse connectives. On the one hand, the two words, “雖然” (suī rán, although) and “但是” (dàn shì, but), which form a word-pair connective, appear in the two discourse units shown in (S1), respectively. These two units demonstrate a *Comparison* relation. On the other hand, “雖然” (suī rán, although) and “但是” (dàn shì, but) can appear individually as single-word connectives shown in (S2)-(S6). The two discourse units have different discourse relations when the single-word connectives appear at different positions, i.e., (S2): *Comparison*, (S3): *Comparison*, (S4): *Expansion*, (S5): *Comparison*, and (S6): *Expansion*. Furthermore, the short word “而” (ér) can be an individual connective, which is interpreted as “而且” (and), “然而” (but), or “因而” (thus), and serves as functions of *Expansion*, *Comparison*, and *Contingency*, respectively. In addition, it can be linked with “雖然” (suī rán, although) and “因為” (yīn wèi, because) to be word-pair connectives, which are interpreted as *Comparison* and *Contingency* functions in (S7) and (S8), respectively. These examples demonstrate word-pair connectives composed of a same word and other words may have different discourse functions, so does the same single-word connective at different positions.

- (S1) 雖然湯姆很聰明，但是他並不用功。(Although Tom is smart, he doesn't study hard.)  
 (S2) 雖然湯姆很聰明，他並不用功。(Although Tom is smart, he doesn't study hard.)  
 (S3) 他流很多汗，雖然才走幾哩路。(He sweated a lot, although he went only a few miles.)  
 (S4) 我會好好閱讀，雖然我真的覺得蜘蛛好可怕。(I'll read, even if I really feel spider terrible.)  
 (S5) 湯姆很聰明，但是他並不用功。(Tom is smart, but he doesn't study hard.)  
 (S6) 但是在巴黎，他放棄了學醫。(But in Paris, he gave up studying medicine.)  
 (S7) 雖然你不說，而我一聞就知道。(Although you did not say, I knew that smell.)  
 (S8) 他因為晚回家，而被媽媽罵了。(Because he came home late, he was scolded by his mother.)

In this paper, we investigate special features of Chinese discourse connectives and apply the results to discourse relation labeling. A semi-supervised learning algorithm is proposed to estimate the probability distribution of the discourse functions of each connective. We address the issue of ambiguity between multiple discourse relations of Chinese connectives. The ambiguity between discourse and non-discourse usages is not our focus in this paper. This paper is organized as follows. Section 2 analyses the types of Chinese connectives and their forward/backward linking properties. Section 3 presents a semi-supervised method to deal with the probability distributions of discourse functions of Chinese connectives and discourse relation labeling. The experimental results are shown and discussed. In Section 4, we further introduce the discourse relation labeler to annotate 302,293 unlabeled sentences and analyze the linguistic phenomena of discourse connectives. We conclude this work in Section 5.

## 2 Types of Discourse Connectives

From the surface form, there are three kinds of linking elements in Chinese (Li and Thompson, 1981): forward-linking elements, backward-linking elements, and couple-linking elements. Discourse connectives are such kinds of linking elements. A discourse unit containing a forward-linking (backward-linking) element is linked with its next (previous) discourse unit. A couple-linking element is a pair of words that exist in two discourse units (Chen, 1994).

Figure 1 shows connectives and their linking direction. The word-pair connective “雖然...但是” (suī rán...dàn shì, although...but) in (S1) is a couple-linking element. A single-word connective may function as a forward-linking element and/or a backward-linking element. It may be a word appearing in a word-pair connective, e.g., “雖然” (suī rán, although), or a word existing individually, e.g., “以及” (yǐ jí, and). A single-word connective which is the first (the second) word of a word-pair connective may function as a forward-linking (backward-linking) element. The single-word connective “雖然” (suī rán, although) in (S2) is a typical example. It keeps the major discourse function, i.e., *Comparison*, of the word-pair connective that it belongs to when it appears in the first discourse unit. In contrast, it may become ambiguous when its position is reversed from the first to the second (i.e., S3 and S4). It may link to the previous or the next discourse units. S5 and S6 have the similar behaviors. The single-word “但是” (dàn shì, but) in (S5) shows a backward-linking. In (S6), it is shifted to the first position and becomes ambiguous. It may be linked to the previous, or to the next discourse units. The correct interpretation depends on the context. These phenomena show a single-word connective may have different senses when it is not at its original position.

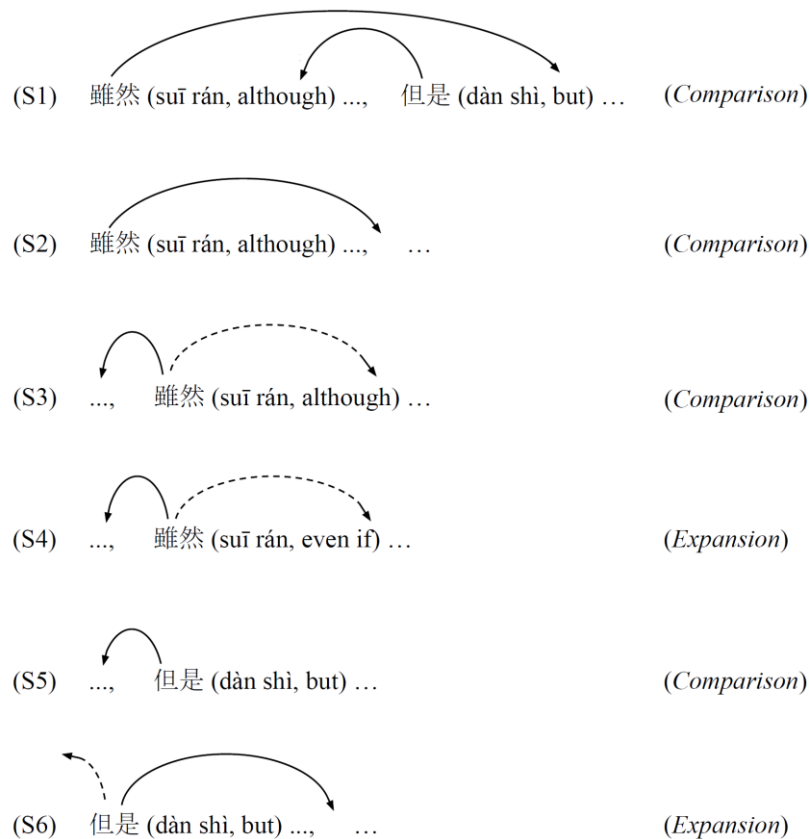


Figure 1: Examples for forward linking and backward linking.

In this study, we collect 808 discourse connectives based on Cheng and Tian (1989), Cheng (2006), and Lu (2007). The discourse connective lexicon contains 319 single-word and 489 word-pair connectives. Initially, each connective is associated with only one discourse function manually by linguists.

For example, the word-pair connective, “雖然...但是” (suī rán...dàn shì, although...but), is assigned a *Comparison* function. The assignment is one-to-one mapping, thus it cannot capture the complete discourse functions of Chinese connectives. Table 1 shows an overview of the discourse connective lexicon. In this lexicon, *Expansion* is the majority, and *Comparison* is the minority. The percentages of *Contingency* and *Expansion* are close. *Temporal* is the third largest discourse function. Intuitively, the discourse connective lexicon cannot cover all their senses. To learn the probability distribution of the discourse functions of a connective needs a large-scale discourse corpus. Compared with RST-DT (Carlson et al., 2002) and PDTB (Prasad et al., 2008), Chinese discourse corpora are not publicly available (Zhou and Xue, 2012; Huang and Chen, 2012b).

Discourse Function	Number of Connectives	Examples of Single-Word and Word-Pair Discourse Connectives
Temporal	151 (18.69%)	接著 (jiē zhe, then), 最初...現在 (zui chū...xiàn zài, first...now)
Contingency	261 (32.30%)	因為 (yīn wèi, because), 如...則 (rú...zé, if ... then)
Comparison	87 (10.77%)	即使 (jí shǐ, even if), 儘管...但 (jìn guǎn...dàn, although...but)
Expansion	309 (38.24%)	另外 (lìng wài, besides), 不僅...而且 (bù jǐn...ér qiě, not only...but also)

Table 1: A Chinese discourse connective lexicon.

### 3 Learning Discourse Functions of Connectives

This section proposes a semi-supervised learning method to learn the interpretation of discourse connectives from an incomplete and sparse dataset.

#### 3.1 A Semi-Supervised Learning Algorithm

Given a pair of discourse units  $ds_1$  and  $ds_2$  containing an explicit connective  $c$ , a discourse relation classifier  $drc$  aims at selecting a relation  $r$  from the set  $\{Temporal, Contingency, Comparison, Expansion\}$  to illustrate how  $ds_1$  and  $ds_2$  cohere to each other. The connective  $c$  may be a word-pair  $c_1...c_2$ , where  $c_1$  and  $c_2$  appear in  $ds_1$  and  $ds_2$ , respectively. It may be a single word appearing in  $ds_1$  or  $ds_2$ . Each discourse unit is mapped into a representation. Various features from different linguistic levels have been explored in the related work (Huang and Chen, 2011; Huang and Chen, 2012a; Zhou et al, 2011; Zhou et al., 2012). We adopt some of their features shown as follows. Here we focus in particular on the probability distributions of the discourse functions and the positions of connectives.

**Length.** This feature includes the word counts of  $ds_1$  and  $ds_2$ .

**Punctuation.** The punctuation at the end of  $ds_2$  is regarded as a feature. The possible punctuation includes a full stop, a question mark, or an exclamation mark. The punctuation at the end of  $ds_1$  is dropped from the features because it is always a comma.

**Words.** The bags of words in  $ds_1$  and  $ds_2$  are considered.

**Hypernym.** The bags of hypernyms of the words in  $ds_1$  and  $ds_2$  are considered. A Chinese thesaurus, Tongyici Cilin<sup>1</sup>, is consulted. The categorization scheme at the fourth level is adopted.

**Shared Word.** The number of words shared in  $ds_1$  and  $ds_2$  is considered as a feature.

**Collocated Word.** Collocated words are word pairs mined from the training set. The first and the second words of a pair come from  $ds_1$  and  $ds_2$ , respectively.

**POS.** The bags of parts of speech in  $ds_1$  and  $ds_2$  are considered.

**Polarity.** Polarity and discourse relation may be related (Huang et al., 2013; Zhou, et al., 2011). For example, a *Comparison* relation implies its two discourse units are contrasting, and some contrasts are presented with different polarities. We estimate the polarity of  $ds_1$  and  $ds_2$  by a lexicon-based approach. The polarity score and the existence of negation are taken as features.

**Discourse Connective.** A discourse connective  $c$  is represented as a probability distribution of discourse functions denoted by a quadruple  $(P_{(c,temporal)}, P_{(c,contingency)}, P_{(c,comparison)}, P_{(c,expansion)})$ , where  $P_{(c,temporal)}$ ,  $P_{(c,contingency)}$ ,  $P_{(c,comparison)}$ , and  $P_{(c,expansion)}$  indicate the probabilities of the four discourse functions of  $c$ , such that  $P_{(c,temporal)}+P_{(c,contingency)}+P_{(c,comparison)}+P_{(c,expansion)}=1$ . Section 3.3 shows how we assign the probabilities to each connective in different experimental settings.

**Position.** The linguistic phenomena discussed in Section 2 show a single-word connective at different position may play different discourse function. Thus, the position of  $c$  is considered as a feature.

<sup>1</sup> <http://ir.hit.edu.cn/>

Because the number of Chinese connectives is large (e.g., 808 Chinese connectives in our lexicon) and the large-scale labeled Chinese discourse corpus is not available, how to learn the probability distribution is a challenging issue. This paper proposes a semi-supervised learning method as follows. Its pseudo code is shown in Algorithm 1.

- (1) Train a 4-way discourse relation classifier  $drc$  with the training set and LIBSVM (Chang and Lin, 2011).
- (2) Initialize probability distributions of unknown connectives in the test set (see experiments).
- (3) Use  $drc$  to label all the instances in the test set.
- (4) Compute the new probability distribution of discourse functions of each connective based on the labeled results in the current run. Maximum likelihood estimation is adopted.
- (5) Repeat (3) and (4) until the number of label changes between two successive runs is below 1%.

---

**Algorithm 1.** Probability Estimation for the Discourse Functions of Connectives

---

**Input:**

$D=\{Temporal, Contingency, Comparison, Expansion\}$ : a set of discourse relations and discourse functions for argument pairs and discourse connectives,

$C=\{c_1, c_2, \dots, c_n\}$ : a set of  $n$  discourse connectives,

$S=\{s_1, s_2, \dots, s_p\}$ : a set of  $p$  labeled argument-pairs  $[sa_1, sa_2]$  containing connective  $c \in CS \subseteq C$ , each with a label  $d \in D$ , where  $CS$  is a set of connectives appearing in  $S$ ,

$T=\{t_1, t_2, \dots, t_q\}$ : a set of  $q$  unlabeled argument-pairs  $[ta_1, ta_2]$  containing connective  $c \in CT \subseteq C$ , where  $CT$  is a set of connectives appearing in  $T$ .

**Output:**

$Q=\{q_1, q_2, \dots, q_n\}$ : a probability distribution  $q_i$  for connective  $c_i \in C$ .

**Method:**

1. Initialization

- 1) Train a classifier  $drc$  using  $S$ .
- 2) Initialize the probability distribution with equal weight, (0.25, 0.25, 0.25, 0.25), for connective  $c \in CT-CS$ , and build  $Q^{(0)}$ .
- 3)  $i \leftarrow 0$

2. Relation labeling

For each  $t \in T$ , estimate the probabilities of four discourse relations,  $P_{(t,temporal)}$ ,  $P_{(t,contingency)}$ ,  $P_{(t,comparison)}$ , and  $P_{(t,expansion)}$ , using the classifier  $drc$  with  $Q^{(i)}$ .

3. Updating the probability distribution

- 1) For each  $c \in C$ , compute the average probability of each discourse relation among the argument-pairs containing  $c$  in  $T$ :

$P_{(c,temporal)} \leftarrow$  Average of  $P_{(t,temporal)}$  for all  $t$  containing  $c$  in  $T$ .

$P_{(c,contingency)} \leftarrow$  Average of  $P_{(t,contingency)}$  for all  $t$  containing  $c$  in  $T$ .

$P_{(c,comparison)} \leftarrow$  Average of  $P_{(t,comparison)}$  for all  $t$  containing  $c$  in  $T$ .

$P_{(c,expansion)} \leftarrow$  Average of  $P_{(t,expansion)}$  for all  $t$  containing  $c$  in  $T$ .

- 2) Form a new  $Q^{(i+1)}$

- 3)  $i \leftarrow i+1$

4. Repeat steps 2-3 until the ratio of the number of label changes by previous and current runs is less than 1%.

5.  $Q \leftarrow Q^{(i)}$
- 

### 3.2 Experimental Setup

For the corpus study of discourse connectives and discourse relations, we refer to a public available Chinese Web POS tagged corpus (Yu et al., 2012). This Chinese POS-tagged corpus is developed based on the ClueWeb09 dataset (CMU, 2009), where Chinese material is the second largest. To capture the discourse functions of individual connectives more accurately, the following three criteria are used to sample sentences:

1. A sentence should contain only two clauses.
2. A sentence should contain exact one discourse connective.

3. The lengths of both clauses in a sentence are no more than 20 Chinese characters.

Total 7,601 sentences composed of two discourse units linked by a connective are sampled from a public available Chinese Web POS tagged corpus (Yu et al., 2012). Each sentence is annotated with a most likely discourse relation selected from  $\{Comparison, Contingency, Comparison, Expansion\}$  by three annotators guided by an instruction manual. The majority is taken as the ground truth. A mentor is involved to make a final decision for the tie conditions. The inter-agreement among the annotators is 0.41 in Fleiss’ Kappa values, which is a moderate agreement. The discourse category with the lowest inter-annotation agreement is *Temporal*, which annotators usually confuse with *Expansion*. It shows the difficulty to distinguish *Temporal* and *Expansion* even by human. Table 2 shows the statistics of the corpus. More than 50% of pairs are annotated with *Expansion* relation. The second largest group is *Contingency* relation. The percentages of *Temporal* and *Comparison* relations are near. Only 359 connectives appear in the corpus. That reflects the incompleteness issue.

Discourse Relation	# Instances	Percentage
Temporal	846	11.13%
Contingency	1,594	20.97%
Comparison	926	12.18%
Expansion	4,235	55.72%

Table 2: Statistics of the experimental discourse corpus.

This Chinese discourse corpus is used for training and testing. We set up the experiments to simulate the scenario of estimating the probability distributions of discourse functions of the unknown connectives based on the information in the training set. We evaluate the experimental results by 5-fold cross-validation. To ensure the discourse connectives appearing in the test set are mutual exclusive of those connectives in the training set, we split the discourse connectives into 5 mutual exclusive sets and split all the 7,601 sentences into 5 folds according to the 5 sets of discourse connectives.

The kernel of our SVM classifier is the radial basis function. The two parameters, cost  $c$  and gamma  $g$ , are optimized by the grid-search algorithm within the range  $c \in \{2^{-5}, 2^{-3}, 2^{-1}, \dots, 2^{15}\}$  and  $g \in \{2^{-15}, 2^{-13}, 2^{-11}, \dots, 2^3\}$ .

### 3.3 Results and Discussions

To demonstrate the performance of our proposed semi-supervised learning methods, the following five models are experimented and compared.

- M0: Label the relation between two discourse units linked by a connective  $c$  based on the  $c$ ’s discourse function defined in the connective lexicon. M0 is considered as a baseline model.
- M1: Train a 4-way discourse relation classifier  $drc$  with the training set, then initialize the function probability distributions of the unknown connectives to  $(0.25, 0.25, 0.25, 0.25)$ , and finally label all the pairs of discourse units by the classifier  $drc$ . M1 is a supervised-learning method.
- M2: M2 model is similar to M1 model except that the probability distribution  $(P_{(c,temporal)}, P_{(c,contingency)}, P_{(c,comparison)}, P_{(c,expansion)})$  of an unknown connective is initialized based on its setting in the connective lexicon. The probability of the unique function is set to 1, and the others are set to 0.
- M3: M3 is a semi-supervised learning method. In testing, the function probability distributions of the unknown connectives are initialized to  $(0.25, 0.25, 0.25, 0.25)$ . Discourse relation labeling and probability distribution updating are done iteratively. Finally, all the test instances are labeled, and probability distributions of discourse functions are learned for all test connectives.
- M4: M4 is similar to M3 except that the initial probability distributions are set based on the connective lexicon.

Table 3 compares the performances of these five models. The average tendency is  $M4 > M3 > M2 > M1 > M0$ . It shows the proposed two semi-supervised learning methods are significantly better than the baseline model M0 and the two supervised-learning methods M1 and M2 at  $p=0.001$ . The best model is M4, but the performance differences between M3 and M4 are not significant. It demonstrates that both the two initial assignments, i.e., equal-weight assignment and lexicon-based

assignment, are effective. If a connective is not listed in the lexicon due to its coverage, we can still derive its probability distribution starting from the equal-weight approach.

We further examine the individual performance of each discourse relation. Comparing M1 and M3, the semi-supervised classifier (M3) outperforms the supervised classifier (M1) in all three metrics in all the four relations except recall and F-score in the *Temporal* relation. Because more than one half of the pairs of discourse units annotated with *Temporal* relation whose discourse connectives have *Expansion* function in the connective lexicon, some discourse-units of *Temporal* relation are misclassified as *Expansion* relation. That is why the recall is dropped by 8.22% in M3. The precisions of all the four relations are increased. In particular, the precisions of *Temporal*, *Contingency*, and *Comparison* gain more than 10%. The overall F-score is increased 6.61%.

Moreover, M4 is better than M2 in F-score for all the relations. In particular, the precisions of *Temporal*, *Contingency*, and *Comparison* recognition by M4 are greatly increased. In other words, the boosting algorithm tends to correct those instances that are originally misclassified into the *Expansion* relation. The t-test also confirms M4 has a significant improvement over M2 at  $p=0.001$ .

The semi-supervised algorithm learns the probability distributions of discourse functions of the unknown connectives from the test instances, so that their size may affect the performance. Figure 2 analyzes how the number of test instances of a connective affects the performance. Each point  $(x, y)$  in this figure denote a connective, where  $x$  is its total occurrences in the test set, and  $y$  is its F-score in Figure 2(a) and its precision/recall in Figure 2(b). We can find (1) many connectives have good performance, (2) connectives containing more test instances demonstrate better performance, and (3) connectives containing fewer instances are sensitive to the evaluation. We treat the probability distribution of discourse functions of each connective as a vector of four real numbers and compute the cosine similarity among the distributions of connectives derived by the connective lexicon, human annotators, and our best model M4. When the 114 connectives containing more than 10 instances are counted, the average cosine similarity between our model and human is 0.940, and the average cosine similarity between the connective lexicon and human is 0.767.

Metric	Model	Temporal	Contingency	Comparison	Expansion	Average
Precision	M0	0.3933	0.7124	0.5092	0.7364	0.6656
	M1	0.5618	0.6005	0.5982	0.7147	0.6595
	M2	0.5024	0.7038	0.5332	0.7529	0.6879
	M3	0.6682	0.7652	0.7749	0.7254	0.7334
	M4	<b>0.6708</b>	<b>0.7773</b>	<b>0.7869</b>	<b>0.7373</b>	<b>0.7344</b>
Recall	M0	0.3757	<b>0.6014</b>	<b>0.6588</b>	0.7389	0.6600
	M1	<b>0.5371</b>	0.5098	0.4154	0.8114	0.6694
	M2	0.4808	0.5808	0.6207	0.7578	0.6731
	M3	0.4549	0.5387	0.5065	<b>0.9015</b>	0.7276
	M4	0.4480	0.5803	0.5821	0.8985	<b>0.7299</b>
F-score	M0	0.3843	0.6522	0.5744	0.7376	0.6606
	M1	<b>0.5492</b>	0.5515	0.4903	0.7600	0.6644
	M2	0.4913	0.6364	0.5736	0.7553	0.6805
	M3	0.5413	0.6323	0.6126	0.8039	0.7305
	M4	0.5372	<b>0.6645</b>	<b>0.6691</b>	<b>0.8099</b>	<b>0.7322</b>

Table 3: Performance comparisons among models.

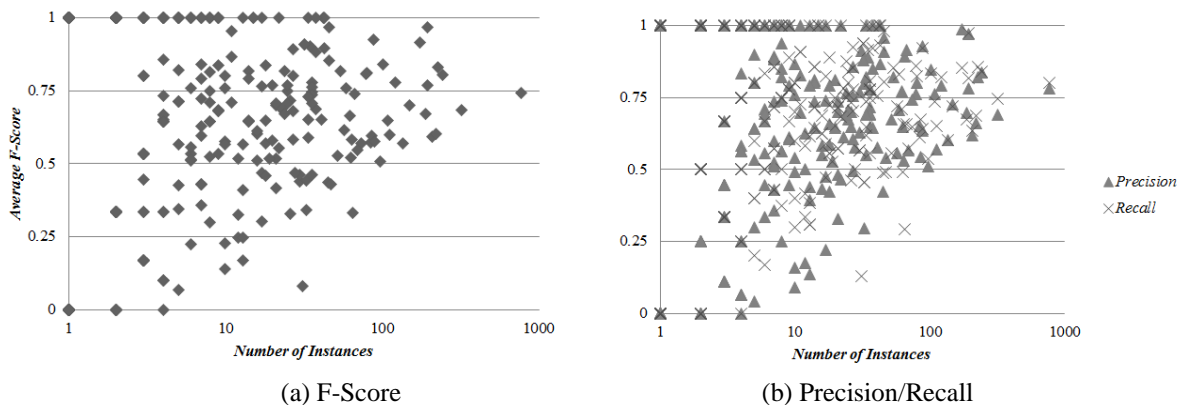


Figure 2: Effects of the number of test instances for each connective on relation labeling.

## 4 Further Analyses on a Big Dataset

We further apply the best model (M4) to predict the probability distributions of discourse functions of connectives on a big dataset. For each discourse connective  $c$ , up to 500 sentences composed of two discourse units linked by  $c$  are randomly selected from the Chinese Web POS tagged corpus (Yu et al., 2012). The limitation of 500 is set to reduce the imbalance among the discourse connectives. Some connectives appear quite often in the dataset, e.g., the connective “也” (yě, also). Some connectives appear less than 500 times, e.g., “千萬...不然” (qiān wàn...bù rán, must...otherwise) occurs only 212 times. Finally, total 302,293 sentences are extracted and predicted. Because the dataset is very large, it is not easy to evaluate each pair of discourse units. We examine the linguistic phenomena instead. A lexicon of the probability distributions of connectives estimated by M4 is available at <http://nlg.csie.ntu.edu.tw/ntu-discourse/>.

We sort the discourse connectives by the ratios of their largest relations. In this way, the top connectives in this order almost contain one relation. They can be considered to be less ambiguous. The top ten connectives which appear 500 times are shown in Table 4. Note the bracket notation  $[ds_1, ds_2]$  denotes the discourse units where connectives appear. The discourse function defined in the discourse connective lexicon specified in Section 2 is marked in bold. The probabilities of the major discourse function of these connectives are larger than 0.89. The distribution is consistent with the human assignment except the last connective “除非...不然” (chú fēi...bù rán, unless...otherwise), which is assigned to *Contingency* in the lexicon. This connective denotes a negated cause-effect relation between  $ds_1$  and  $ds_2$  in which  $ds_2$  is the effect when  $ds_1$  is not satisfied. In such a case,  $ds_1$  and  $ds_2$  show clear contrast, so that it is reasonable to label this connective with a higher probability of the *Comparison* relation. There are two groups of synonyms in the list: (1) “雖然...不過” (suī rán...bú guò, although...but) and “雖然...可是” (suī rán...kě shì, although...but), and (2) “簡言之” (jiǎn yán zhī, in short) and “簡而言之” (jiǎn ér yán zhī, in short). Table 4 shows that synonyms share similar distributions. The cosine similarities of their probability distributions are 0.99996 and 0.99952, respectively.

The probability of each discourse function of each connective  $c$  is the average of the probabilities estimated by the classifier, thus the distributions reported by our model is not completely identical to the empirical distribution. For example, all the instances containing the connective “雖然...不過” (suī rán...bú guò, although...but) are labeled with the major discourse function *Expansion*, but the estimated probability of *Expansion* of this connective is 93.47%.

We also sort the discourse connectives by the ratio of their second largest relations. In this manner, the top connectives in this order may have two major discourse functions. In other words, they are ambiguous. Table 5 shows the top ten estimated ambiguous discourse connectives. It is interesting that *Expansion* is one of the two major discourse functions, and the other one shown in bold is the discourse function defined in the connective lexicon. The discourse connectives “緊接著” (jǐn jiē zhe, then), “現在” (xiàn zài, now), “未來” (wèi lái, in the future), and “終於” (zhōng yú, finally), which are defined to have *Temporal* function in the lexicon, frequently occur in the discourse units with *Expansion* relation. The estimated distribution of the connective “而” (ér, and; but; thus) is consistent with the human interpretation, i.e., it has multiple discourse functions.

Chinese single-word connectives are usually put together with other words to form word-pair connectives. Tables 6 and 7 show examples for “雖然” (suī rán, although) and “所以” (suǒ yǐ, so),

Discourse Connectives $[ds_1, ds_2]$	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[簡言之, ...] ([in short, ...])	2.78	2.08	1.67	<b>93.47</b>
[雖然, 不過] ([although, but])	0.77	1.80	<b>92.70</b>	4.74
[換言之, ...] ([in other words, ...])	3.63	2.82	1.53	<b>92.02</b>
[雖然, 可是] ([although, but])	0.93	2.11	<b>91.58</b>	5.37
[由於, 因此] ([since, therefore])	1.41	<b>91.07</b>	0.97	6.55
[說到底, ...] ([after all, ...])	3.17	3.95	2.97	<b>89.91</b>
[..., 說到底] ([..., after all])	3.13	4.34	2.84	<b>89.69</b>
[簡而言之, ...] ([in short, ...])	5.07	3.20	2.25	<b>89.48</b>
[或是, 或是] ([or, or])	3.94	4.51	2.16	<b>89.39</b>
[除非, 不然] ([unless, otherwise])	1.04	<b>3.71</b>	89.33	5.93

Table 4: Top 10 less-ambiguous connectives estimated by using a big dataset.



Discourse Connectives [ds1, ds2]	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[緊接著, ...] ([then, ...])	<b>48.71</b>	5.12	1.70	44.46
[..., 即使] ([..., even though])	3.93	5.23	<b>46.48</b>	44.36
[現在, ...] ([now, ...])	<b>44.31</b>	7.42	3.42	44.85
[..., 雖然] ([..., although])	3.60	3.68	<b>44.17</b>	48.55
[以便, ...] ([so that, ...])	3.96	<b>49.83</b>	2.05	44.16
[目前, 未來] ([now, in the future])	<b>47.05</b>	6.41	3.10	43.44
[只有, 才] ([only, then])	4.34	<b>43.30</b>	9.33	43.03
[未來, ...] ([in the future, ...])	<b>48.21</b>	6.15	2.85	42.79
[..., 而] ([..., and; but; thus])	3.72	6.13	<b>42.78</b>	47.37
[..., 終於] ([..., finally])	<b>42.39</b>	6.13	2.99	48.49

Table 5: Some ambiguous connectives estimated by using a big dataset.

respectively. The former is often connected with a word in the second discourse unit to form a couple-linking, while the latter is connected with a word in the first one. We can find word-pair connectives are less ambiguous than single-word connectives in different probabilities. The former (“雖然”, suī rán, although) tends to have *Comparison* function. When the word-pair connectives are shorten to single-word connectives, the probability to have *Comparison* function becomes lower. The connective “雖然” (suī rán, although) in the first argument still has probability 0.7639 to have *Comparison* function. When “雖然” (suī rán, although) is moved to the second argument, the probability to serve as *Comparison* function is decreased to 0.4417, which is even lower than that of *Expansion* function. It shows that couple-linking elements provide strong clue to determine discourse relation. Besides, a single-word connective has some tendency to function as either forward linking or backward linking. For example, “雖然” (suī rán, although) is a forward-linking element. Normally, it will link the first discourse unit containing it with the second one. When it appears in the second discourse unit, it becomes ambiguous. The connectives containing “所以” (suǒ yǐ, so) have the similar effects. It tends to be a backward linking element, so its companion appears in the first discourse unit. Its probability to have *Contingency* function decreases from a word-pair connective to a single-word connective. When it appears in the first discourse unit, it may link to the previous sentence at the inter-sentential level.

Some Chinese short words like “而” (ér) is often a part of word-pair connectives. Table 8 shows 10 words which are often connected with “而” (ér) to form word-pair connectives. The word-pair connectives tend to have one major function. When the word-pair connective is “abbreviated” to a single-

Discourse Connectives [ds1, ds2]	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[雖然, 不過] ([although, but])	0.77	1.80	<b>92.70</b>	4.74
[雖然, 可是] ([although, but])	0.93	2.11	<b>91.58</b>	5.37
[雖然, 然而] ([while, however])	1.04	2.03	<b>90.76</b>	6.17
[雖然, 但是] ([although, but])	1.14	2.62	<b>88.49</b>	7.74
[雖然, 但] ([although, but])	1.48	2.89	<b>87.54</b>	8.09
[雖然, 還] ([although, still])	2.70	3.43	<b>85.20</b>	8.68
[雖然, 仍] ([although, still])	3.06	4.10	<b>81.03</b>	11.81
[雖然, 而] ([although, while])	2.86	5.09	<b>79.23</b>	12.82
[雖然, 仍然] ([although, still])	3.68	5.70	<b>77.23</b>	13.39
[雖然, 還是] ([although, still])	3.51	8.54	<b>75.26</b>	12.69
[雖然, 卻] ([although, still])	4.24	3.71	<b>74.58</b>	17.47
[雖然, ...] ([although, ...])	3.46	5.28	<b>76.39</b>	14.87
[..., 雖然] ([..., although])	3.60	3.68	<b>44.17</b>	<b>48.55</b>

Table 6: Effects of single-word and word-pair connectives containing “雖然” (suī rán, although).

Discourse Connectives [ds1, ds2]	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[由於, 所以] ([because, so])	1.64	<b>85.25</b>	1.77	11.35
[因, 所以] ([because, so])	2.26	<b>83.20</b>	1.82	12.72
[因為, 所以] ([because, so])	2.69	<b>78.03</b>	2.35	16.93
[既然, 所以] ([since, so])	1.68	<b>67.32</b>	6.37	24.63
[..., 所以] ([..., so])	2.82	<b>50.67</b>	5.29	<b>41.22</b>
[所以, ...] ([so, ...])	5.71	<b>50.61</b>	2.50	<b>41.18</b>

Table 7: Effects of single-word and word-pair connectives containing “所以” (so).

word connective, it becomes ambiguous. The discourse function depends on which word-pair connective it is mapped. The determination relies on contextual information.

Table 9 further shows the effects of positions of single-word connectives. The major discourse function of the first 7 sets of connectives is changed when the connectives are shifted from the first discourse unit to the second one. In contrast, the last 3 sets of connectives keep their major discourse function no matter whether they are placed in the first or the second discourse unit. The only difference is the probability to serve as the major discourse function is changed. For example, the probability of the connective “只不過” (zhǐ bú guò, only; just; merely) to have *Comparison* function is increased from 0.6920 to 0.8501 when it is shifted from the first discourse unit to the second one.

Discourse Connectives [ds1, ds2]	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[不只, 而] ([not only, but])	2.19	4.13	4.92	<b>88.76</b>
[不僅, 而] ([not only, but])	2.41	4.56	10.13	<b>82.89</b>
[不但, 而] ([not only, but])	3.20	5.14	10.55	<b>81.11</b>
[既然, 而] ([since, but])	3.99	13.87	13.42	<b>68.72</b>
[固然, 而] ([of course, while])	1.16	2.76	<b>80.82</b>	15.24
[雖然, 而] ([although, while])	2.86	5.09	<b>79.23</b>	12.82
[儘管, 而] ([although, while])	2.76	43.61	<b>79.16</b>	13.71
[由於, 而] ([because, so])	2.02	<b>79.01</b>	2.16	16.81
[因, 而] ([because, so])	3.21	<b>71.03</b>	2.28	23.49
[因為, 而] ([because, so])	3.11	<b>49.12</b>	7.52	40.26
[..., 而] ([..., and; but; thus])	3.71	6.13	<b>42.78</b>	<b>47.37</b>
[而, ...] ([and; but; thus, ...])	5.47	8.55	17.00	<b>68.98</b>

Table 8: Effects of single-word and word-pair connectives containing “而” (and, but, so).

Discourse Connectives [ds1, ds2]	Temporal (%)	Contingency (%)	Comparison (%)	Expansion (%)
[因而, ...] ([therefore, ...])	6.26	<b>64.30</b>	1.66	27.77
[..., 因而] ([..., therefore])	3.54	28.32	5.15	<b>62.99</b>
[只要, ...] ([as long as, ...])	2.68	<b>66.02</b>	5.33	25.98
[..., 只要] ([..., as long as])	2.57	5.49	4.23	<b>87.71</b>
[假如, ...] ([if, ...])	3.51	<b>57.15</b>	7.47	31.87
[..., 假如] ([..., if])	3.31	5.21	5.33	<b>86.16</b>
[不過, ...] ([however, ...])	8.17	9.20	23.12	<b>59.51</b>
[..., 不過] ([..., however])	2.26	2.39	<b>80.97</b>	14.38
[但是, ...] ([but, ...])	8.56	7.72	20.87	<b>62.86</b>
[..., 但是] ([..., but])	2.32	2.90	<b>75.76</b>	19.02
[即使, ...] ([even though, ...])	3.55	5.04	<b>75.65</b>	15.75
[..., 即使] ([..., even though])	3.93	5.23	<b>46.48</b>	<b>44.36</b>
[現在, ...] ([now, ...])	<b>44.31</b>	7.42	3.42	<b>44.85</b>
[..., 現在] ([..., now])	8.03	2.88	3.60	<b>85.49</b>
[且, ...] ([and, ...])	7.14	8.43	3.14	<b>81.29</b>
[..., 且] ([..., and])	4.62	3.79	2.38	<b>89.22</b>
[以及, ...] ([as well as, ...])	4.83	9.88	2.69	<b>82.60</b>
[..., 以及] ([..., as well as])	4.20	4.29	2.33	<b>89.18</b>
[只不過, ...] ([merely, ...])	3.54	4.76	<b>69.20</b>	22.50
[..., 只不過] ([..., merely])	1.48	2.00	<b>85.01</b>	11.50

Table 9: Effects of positions of single-word connectives.

## 5 Conclusion

In this paper, we address the issue of the ambiguous discourse functions of Chinese connectives in discourse relation labeling and propose a semi-supervised learning method to estimate the probability distribution of discourse functions of connectives. We examine the constructions of Chinese connectives and their effects on the discourse relation recognition. The proposed approach learns the probability distributions of discourse functions of Chinese connectives from a small labeled dataset and a big unlabeled dataset. The results reflect many interesting linguistic phenomena. We compare the ambiguity degrees of single-word and word-pair connectives, and show the effects of the positions of single-word connectives on the discourse functions. The discourse relation recognizer integrating the

probability distributions and contextual information significantly outperforms the approaches without the knowledge.

This methodology can be extended to estimate the probability distribution of discourse functions of connectives on much finer relation categories. In the current experiments, we focus on explicit discourse relation recognition. The 302,293 labeled sentences in Section 4 can be regarded as a training corpus for implicit discourse relation recognition. Those labeled sentences composed of unambiguous connectives will be sampled from the reference corpus for training an implicit discourse relation recognition system. Furthermore, how to employ the learned probability distributions to deal with discourse units containing multiple connectives will be investigated. In the future, we will tell out the discourse connective and non-discourse connective uses of words and explore their interpretations on the discourse relation recognition. Besides, we will make use of the probability distributions to the relation labeling on more than two clauses and further extend the methodology to experiments at the inter-sentence level.

## Acknowledgements

This research was partially supported by Ministry of Science and Technology, Taiwan, under the grants 101-2221-E-002-195-MY3 and 102-2221-E-002-103-MY3, and 2012 Google Research Award. We are also very thankful to the anonymous reviewers for their helpful comments to revise this paper.

## References

- Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Péry-Woodley, Laurent Prévot, Josette Rebeyrolle, Ludovic Tanguy, Marianne Vergez-Couret, and Laure Vieu. 2012. An Empirical Resource for Discovering Cognitive Principles of Discourse Organisation: the ANNODIS Corpus. In *Proceedings of the 18th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 2727-2734, Istanbul, Turkey.
- Nicholas Asher and Alex Lascarides. 1995. Lexical Disambiguation in a Discourse Context. *Journal of Semantics*, 12(1):69-108, Oxford University Press.
- Lynn Carlson, Daniel Marcu, and Mary E. Okurowski. 2002. *RST Discourse Treebank*. Linguistic Data Consortium, Philadelphia.
- Bruno Cartoni, Sandrine Zufferey, and Thomas Meyer. 2013. Annotating the Meaning of Discourse Connectives by Looking at their Translation: The Translation Spotting Technique. *Dialogue and Discourse*, 4(2):65-86.
- Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1-27:27.
- Hsin-Hsi Chen. 1994. The Contextual Analysis of Chinese Sentences with Punctuation Marks. *Literal and Linguistic Computing*, 9(4):281-289.
- Shou-Yi Cheng. 2006. *Corpus-Based Coherence Relation Tagging in Chinese Discourse*. Master Thesis, National Chiao Tung University, Hsinchu, Taiwan.
- Xianghui Cheng and Xiaolin Tian. 1989. *Xian dai Han yu (現代漢語)*, San lian shu dian (三聯書店), Hong Kong.
- CMU 2009. ClueWeb09, <http://lemurproject.org/clue-web09.php/>
- Leon Derczynski and Robert Gaizauskas. 2013. Temporal Signals Help Label Temporal Relations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Volume 2: Short Papers, pages 645-650, Sofia, Bulgaria.
- Jerry R. Hobbs. 1985. On the Coherence and Structure of Discourse, Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University. <http://www.isi.edu/~hobbs/ocsd.pdf>
- Eduard H. Hovy and Elisabeth Maier. 1992. Parsimonious or Profligate: How Many and Which Discourse Structure Relations? No. ISI/RR-93-373. Information Sciences Institute, University of Southern California, Marina del Rey.
- Hen-Hsen Huang and Hsin-Hsi Chen. 2011. Chinese Discourse Relation Recognition. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, pages 1442-1446, Chiang Mai, Thailand.

- Hen-Hsen Huang and Hsin-Hsi Chen. 2012a. Contingency and Comparison Relation Labeling and Structure Prediction in Chinese Sentences. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2012)*, pages 261-269, Seoul, South Korea.
- Hen-Hsen Huang and Hsin-Hsi Chen. 2012b. An Annotation System for Development of Chinese Discourse Corpus. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012) Demonstration Papers*, pages 223-230, Mumbai, India.
- Hen-Hsen Huang, Chi-Hsin Yu, Tai-Wei Chang, Cong-Kai Lin, and Hsin-Hsi Chen. 2013. Analyses of the Association between Discourse Relation and Sentiment Polarity with a Chinese Human-Annotated Corpus. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 70-78, Sofia, Bulgaria.
- Ben Hutchinson. 2004. Acquiring the Meaning of Discourse Markers. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, pages 684-691, Barcelona, Spain.
- Charles N. Li, Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.
- Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2011. Automatically Evaluating Text Coherence Using Discourse Relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, pages 997-1006, Portland, Oregon, USA.
- Shuxiang Lu. 2007. *Eight Hundred Words of The Contemporary Chinese (Xian dai Han yu Ba bai Ci)*. China Social Sciences Press.
- Emily Pitler and Ani Nenkova. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 13-16, Suntec, Singapore.
- Livia Polanyi. 1988. A Formal Model of the Structure of Discourse. *Journal of Pragmatics*, 12(5-6):601-638.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*, pages 2961-2968, Marrakech, Morocco.
- Charlotte Roze, Laurence Danlos, and Philippe Muller. 2010. LEXCONN: a French Lexicon of Discourse Connectives. In *Proceedings of the 8th International Workshop on Multidisciplinary Approaches to Discourse (MAD 2010)*, Moissac.
- Ted J. M. Sanders, Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. Toward a Taxonomy of Coherence Relations. *Discourse Processes*, 15(1):1-35.
- Caroline Sporleder and Alex Lascarides. 2008. Using Automatically Labelled Examples to Classify Rhetorical Relations: A Critical Assessment. *Natural Language Engineering*, 14(3):369-416, Cambridge University Press.
- Fei Wang, Yunfang Wu, and Likun Qiu. 2012. Exploiting Discourse Relations for Sentiment Analysis. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Posters, pages 1311-1320, Mumbai, India.
- Florian Wolf and Edward Gibson. 2005. Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*, 31(2):249-287.
- Chi-Hsin Yu, Yi-jie Tang and Hsin-Hsi Chen. 2012. Development of a Web-scale Chinese Word N-gram Corpus with Parts of Speech Information. In *Proceedings the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 320-324, Istanbul, Turkey.
- Lanjun Zhou, Binyang Li, Wei Gao, Zhongyu Wei and Kam-Fai Wong. 2011. Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 162-171, Edinburgh, UK.
- Lanjun Zhou, Wei Gao, Binyang Li, Zhongyu Wei and Kam-Fai Wong. 2012. Cross-lingual Identification of Ambiguous Discourse Connectives for Resource-Poor Language. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 1409-1418, Mumbai, India.
- Yuping Zhou and Nianwen Xue. 2012. PDTB-style Discourse Annotation of Chinese Text. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, pages 67-77, Jeju Island, Korea.