SoNLP-DP System for ConLL-2016 Chinese Shallow Discourse Parsing

Junhui Li[1] **Fang Kong**[1] **Sheng Li**[2] **Muhua Zhu**[2] **Guodong Zhou**[1]
[1]Natural Language Processing Lab, Soochow University, China
{lijunhui, kongfang, gdzhou}@suda.edu.cn
[2] Alibaba Inc., Hangzhou, China
{lisheng.ls, muhua.zmh}@alibaba-inc.com

Abstract

This paper describes our submission to the CoNLL-2016 shared task (Xue et al., 2016) on end-to-end Chinese shallow discourse parsing. We decompose the end-to-end process into four steps. Firstly, we define a syntactically heuristic algorithm to identify elementary discourse units (EDUs) and further to recognize valid EDU pairs. Secondly, we recognize explicit discourse connectives. Thirdly, we link each explicit connective to valid EDU pairs to obtain explicit discourse relations. For those valid EDU pairs not linked to any explicit connective, they become non-explicit discourse relations.[1] Finally, we assign each discourse relation, either explicit or non-explicit with a discourse sense. Our system is evaluated on the closed track of the CoNLL-2016 shared task and achieves 35.54% and 23.46% in F1-measure on the official test set and blind test set, respectively.

1 Introduction

Shallow discourse parsing maps a piece of text into a set of discourse relations, each of which is composed of a discourse connective, two arguments, and the sense of the discourse connective. Shallow discourse parsing has been drawing more and more attention in recent years due to its importance in deep NLP applications, such as coherence modeling (Barzilay and Lapata, 2005; Lin et al., 2011), event extraction (Li et al., 2012), and statistical machine translation (Tu et al., 2014).

During the past few years, English shallow discourse parsing has dominated the research on discourse parsing, thanks to the availability of Penn Discourse TreeBank (PDTB) (Prasad et al., 2008). As a representative, Lin et al. (2014) decompose the end-to-end PDTB-styled discourse parser into a few components, including a connective classifier, an argument labeler, an explicit sense classifier, and a non-explicit sense classifier. The popularity of English shallow discourse parsing is further fueled by the CoNLL-2015 shared task (Xue et al., 2015). Meanwhile research on Chinese discourse parsing is also carried out smoothly (Zhou and Xue, 2012; Li et al., 2014). As a complement to PDTB annotated on English TreeBank, Chinese Discourse TreeBank (CDTB) (Zhou and Xue, 2012) annotates shallow discourse relations on Chinese TreeBank by using similar framework of PDTB. However, the two languages have many different properties. For example, the non-explicit discourse relations in the training data of CoNLL-2016 shared task dataset account for 54.75% in English while they account for 78.27% in Chinese, indicating the difficulties in Chinese shallow discourse parsing. Second, the two arguments of a Chinese non-explicit discourse relation are more apt to locate in the same sentence. This is verified by the statistics that 56.57% of Chinese non-explicit discourse relations are within one sentence while only 2.55% of English non-explicit discourse relations are. In particular, the English non-explicit discourse relations are usually composed of two consecutive sentences.

This paper describes our submission to the CoNLL-2016 shared task on end-to-end Chinese shallow discourse parsing. A participant system needs to (1) identify all explicit discourse connectives in the text (e.g., continuous connectives "尽管", "另 一 方面", discontinuous one "由于 ... 因此"), (2) identify the spans of text that function as the two arguments (i.e., Arg1 and Arg2) for each discourse connective, and (3) predict the

[1]In this paper, non-explicit discourse relations include discourse relations with type *implicit*, *entrel*, and *altlex*.

Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 78–84,
Berlin, Germany, August 7-12, 2016.

sense of the discourse relations (e.g., *Cause*, *Condition*, *Contrast*). Due to the differences between Chinese and English, our approach to Chinese discourse parsing is very different from the one to English discourse parsing (Lin et al., 2014; Kong et al., 2014). For example, Lin et al. (2014) construct non-explicit discourse relations in English by looking for two consecutive sentences that are not connected to any explicit connective. However, it fails to discover non-explicit discourse relations in which the two arguments locate in one sentence. Alternatively, we decompose the whole process of our Chinese discourse parser into four steps. Firstly, we define a syntactically heuristic algorithm to identify elementary discourse units (EDUs) and further to recognize valid EDU pairs. Secondly, we recognize explicit discourse connectives. Thirdly, we link each explicit connective to valid EDU pairs to obtain explicit discourse relations. For those valid EDU pairs not linked to any explicit connective, they become non-explicit discourse relations. Finally, we assign each discourse relation, either explicit or non-explicit with a discourse sense. Our system is evaluated on the closed track of the CoNLL-2016 shared task and achieves 35.54% and 23.46% in F1-measure on the official test set and blind test set, respectively.

The rest of this paper is organized as follows. Section 2 describes the details of our Chinese shallow discourse parser. In Section 3, we present our experimental results, followed by the conclusion in Section 4.

2 System Architecture

In this section, we first present an overview of our system. Then we describe the details of our components in the end-to-end Chinese discourse parser.

2.1 System Overview

A typical text consists of sentences glued together in a systematic way to form a coherent discourse. In PDTB and CDTB, shallow discourse parsing focuses on shallow discourse relations either lexically grounded in explicit discourse connectives or associated with sentential adjacency. Different from deep discourse parsing, shallow discourse parsing transforms a piece of text into a set of discourse relations between two adjacent or non-adjacent discourse units, instead of connecting the relations hierarchically to one another to form a connected structure in the form of tree or graph.

Specifically, given a piece of text, the end-to-end shallow discourse parser returns a set of discourse relations in the form of a discourse connective (explicit or non-explicit) taking two arguments with a discourse sense. Figure 1 shows the framework of our end-to-end system which consists of six components (i.e., from *A* to *F*). Next, we decompose the process into four steps:

- Firstly, we define a heuristic algorithm to identify elementary discourse units (EDUs) and further to recognize valid EDU pairs. This step includes components of *A* and *B* in Figure 1.
- Secondly, we recognize explicit discourse connectives. This is task of component *C* in Figure 1.
- Thirdly, we link each explicit connective to valid EDU pairs to obtain explicit discourse relations. For those valid EDU pairs not linked to any explicit connective, they become non-explicit discourse relations. This is what component *D* does in Figure 1.
- Finally, we assign each discourse relation, either explicit or non-explicit with a discourse sense. Specifically, we use component *E* to assign sense for explicit discourse relations while using component *F* for non-explicit discourse relations.

2.2 EDU Identification

An EDU is a sequence of words that represents an event, which is usually driven by a VP (a.k.a. verbal phrase) node in parse tree. Given a parse tree, we collect all basic VPs in it. In contrast to a nested VP that is composed of either multiple sub-VPs or a VP and its modifiers, a basic VP is a VP that headed by a non-VP. For example, in Figure 2, VP_2 and VP_4 are basic VPs since VP_2 is headed by VE/无 while VP_4 is headed by VV/通过. In contrast, VP_1 and VP_3 are not basic VPs since they are both headed by basic VPs, i.e., VP_2 and VP_4. For each basic VP, we use the heuristic Algorithm 1 to find its left and right boundary nodes, and thus obtain the word sequence representing the corresponding EDU.

It is easy to find the right boundary node since we always set it as the basic VP node (*line1*). The algorithm initializes the left boundary node as the

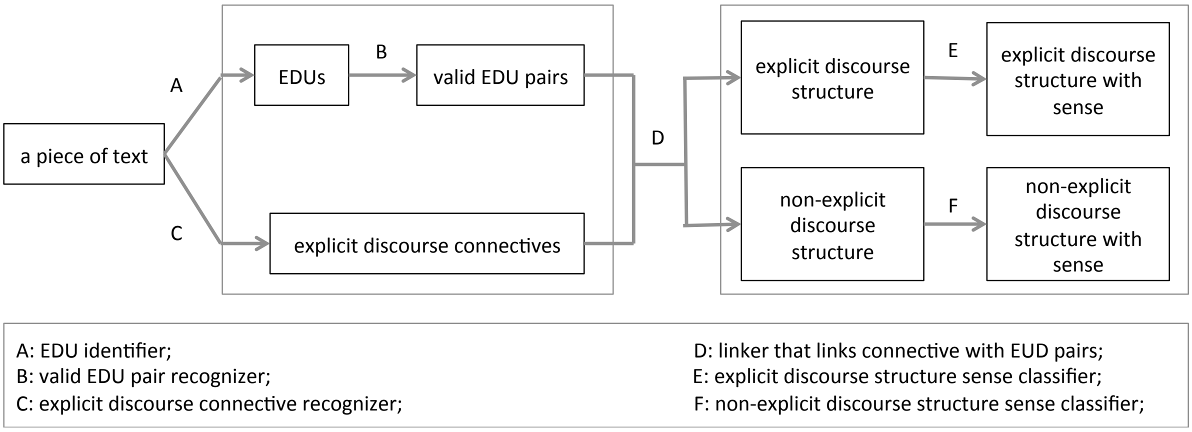

Figure 1: Framework of our end-to-end Chinese shallow discourse parser.

Algorithm 1: Obtaining EDU from a basic VP

Input: parse tree *tree*
basic VP node vp
Output: its corresponding EDU
1. define right boundary node $rbn = vp$;
2. define left boundary node $lbn = vp$;
3. set current node c as vp;
4. **while** $(true)$
5. set node p as c's parent;
6. **if** $(p == null)$ **break**;
7. get p's production rule, say as l_m .. l_1 c r_1 ..r_n, indicating c has m left hand siblings and n right siblings;
8. **for** i from 1 to m
9. **if** l_i is dominated by c
10. $lbn = l_i$;
11. **else**
12. **break**;
13. **if** $i <= m$ **break**;
14. $c = p$;
15. return word sequence from position leftmost of lbn to rightmost of rbn;

basic VP node as well (*line2*). Then it repeatedly update the left boundary node until it finds a proper one. To this end, the algorithm starts by setting the current node c as the basic VP node (*line 3*), and first examine the left siblings from right to left and see if they are dominated by c. It then iteratively moves one level up to the parent of c till it reaches the root of the tree (*line 14*). At each level, it repeatedly updates the left boundary node (*line 10*). Specifically, if there exists a left sibling which is *not* dominated by c, the algorithm stops (*line 12 & 13*). Once both the left and right boundary nodes are found. It uses the leftmost position of the left boundary node and the rightmost position of the right boundary node to obtain the word sequence of the corresponding EDU. For VP_2 and VP_4 in Figure 2, the algorithm will return "第二 天 无 大 新闻" and "种子 选手 均 顺利 通过 第一 轮" as their EDUs, respectively.

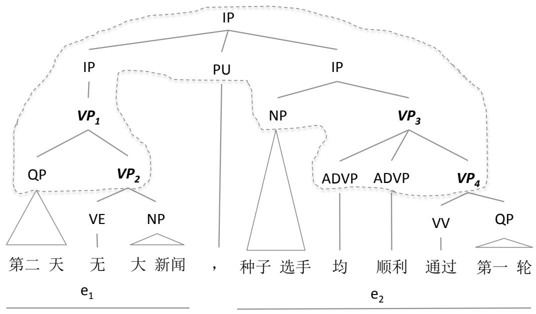

Figure 2: An example of recognizing EDUs.

Note that for two EDUs that occur in one sentence, they satisfy that either their spans have no overlapping at all (e.g., e_1 and e_2 in Figure 2), or one EDU fully covers the other.

2.3 Valid EDU Pair Recognition

A valid EDU pair is two EDUs that have discourse relation, either explicit or non-explicit. We first collect all potential EDU pairs as candidate, and then identify valid ones. In an EDU pair, we presume the first EDU locates on the left side of the second one.

Intra-EDU pair candidates. Intra-EDU pair candidates indicate that the two focusing EDUs locate in one sentence. If a sentence contains two or more EDUs, we enumerate all possible EDU pairs as candidates as long as the pair have no overlapping in position.

Inter-EDU pair candidates. The two EDUs in an inter-EDU pair candidate locate in two sentences. To make the task simple, we only consider such candidates if the two EDUs are in two consecutive sentences. For two consecutive sentences s_1 and s_2, we obtain their corresponding set (es_1 and es_2) of EDUs that are at top level (i.e., an EDU is at top level if it is not covered by another EDU). Then we enumerate all possible EDU pairs by selecting one from es_1 and the other from es_2.

To identify an EDU pair candidate is valid or not, we use tree kernel approach to explore implicitly structured features by directly computing the similarity between two subtrees. Given a parse tree and an EDU pair candidate in it,[2] we first find the lowest ancestor node that fully covers the two EDUs. Then we collect left and right siblings along the path from the lowest ancestor node to each basic VP node. For example, the dash circle in Figure 2 represents the subtree for the EDU pair of e_1 and e_2.

2.4 Explicit Discourse Connective Recognition

Connectives in Chinese are more obscure than those in English. For example, we extract 358 types of connective from the training data. Among them, 193 (or 54%) types of connective occur once while 197 (or 55%) types consist of two or more words. Being worse, 32 (or 9%) types of connective span two or more sentences. Our system keeps 326 (or 91%) types of connective that locate in one sentence as our connective set. That is to say, we ignore those connectives that locate in two or more sentences. The distribution of connective in training data suggests that the connective set is an open set. Given a piece of text, we first use the connective set to collect connective candidates. Then we identify each connective candidate is a functional connective or not. Different from previous work that defines diverse linguistic features, varying from lexical knowledge to syntactic parse trees, we use tree kernel approach to explore implicitly structured features by directly computing the similarity between two subtrees. Given a parse tree and a connective candidate in it, we first find the lowest *IP* node that fully covers the connective. Then we collect left and right siblings along the path from the *IP* node to each connective word. For instance, sentence "由于 新 组建 的 国家队 新 队员 将 占 一半 ， 而 她们 的 技术 水平 尚 待 提高 ， 因此 面临 的 任务 是 艰巨 的 " and

[2]for inter-EDU pair candidate, we manually create a top node and take the parse trees of the two consecutive sentences as children of top node.

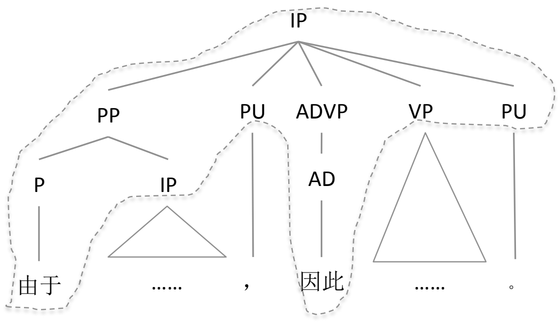

Figure 3: An example of subtree extraction for connective recognition.

a discontinuous connective candidate "由于 ... 因此" in it, we extract a subtree as shown in Figure 3.

2.5 Linking connective with EDU pairs

So far we have recognized both valid EDU pairs and explicit discourse connectives. Our next step is to link a connective to EDU pairs. Note that it is possible for a connective to link to one or more EDU pairs. To decide if a connective and an EDU pair is relevant, we continue to use tree kernel approach. The subtrees extraction algorithm is very similar to that of valid EDU pair recognition. The algorithm first finds the lowest ancestor node that covers the two EDUs and the connective. Then it collects left and right siblings along the path from the lowest ancestor node to connective word, and to the two basic VP nodes, respectively. For instance, in sentence "由于 新 组建 的 国家队 新 队员 将 占 一半 ， 而 她们 的 技术 水平 尚 待 提高 ， 因此 面临 的 任务 是 艰巨 的 ", we are about to predict if the connection exist between a discontinuous connective "由于 ... 因此" and an EDU pair colored in blue and green in Figure 4. To this end, the subtree extraction algorithm first looks for their lowest ancestor, i.e., the top *IP* in Figure 4, then the algorithm collect all siblings along the paths from the lowest ancestor node (i.e., *IP*) to each connective word (i.e., *P* and *ADVP*), and to the two basic VPs (i.e., the two colored VPs). Figure 4 also shows the extracted subtree.

Explicit discourse relations. If one or more valid EDU pairs are predicted to have connection to a connective,[3] we construct an explicit dis-

[3]If none EDU pair is predicted to have connection to a connective, we take the pair with the highest probability as the one linking to the connective.

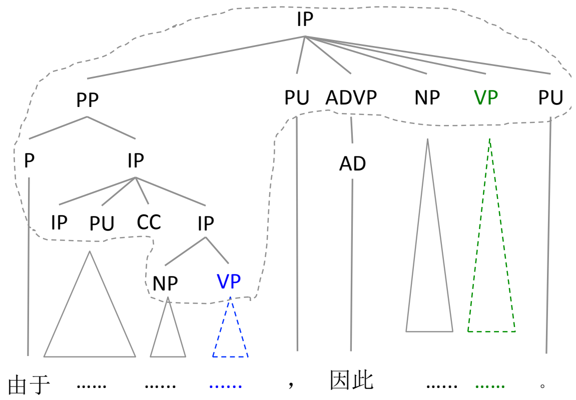

Figure 4: An example of subtree extraction for linking a connective with an EDU pair.

course relation by merging all the first EDUs of the EDU pairs as Arg1 of the connective, and merging all the second EDUs of the EDU pairs as Arg2.

Non-explicit discourse relations. If a valid EDU pair is not linked to any explicit connective, we construct a non-explicit discourse relation by regarding the first EDU as Arg1 and the second as Arg2.

2.6 Sense Classification for Explicit discourse relations

Once an explicit discourse relation is identified, the sense classifier is used to predict its sense. Due to the fact the connective themselves are strong hint for their sense, we follow (Lin et al., 2014) to define a few lexical features to train a sense classifier: the connective words themselves, their part-of-speeches and the previous words of each connective word.

2.7 Sense Classification for Non-explicit discourse relations

Due to the absence of discourse connectives, sense prediction for non-explicit discourse relations is more difficult. Following the work of Kong et al. (2015) on non-explicit sense classification in English, we define the following groups of features:

- **First three words of Arg1/Arg2:** This set of features include the first three words in Arg1 and Arg2.

- **Production rules:** According to Lin et al. (2009), the syntactic structure of one argument may constrain the relation type and the syntactic structure of the other argument. We extract production rules from training data with frequency larger than 5 times. Then for each production rule *pr*, we add features *pr-in-arg1=1*, *pr-in-arg2=1*, *pr-in-arg1arg2=1* if it occurs in Arg1, Arg2, and both, respectively.

- **Dependency rules:** Similar to the above features of production rules, three sets of features *dr-in-arg1=1*, *dr-in-arg2=1*, *dr-in-arg1arg2=1* if it occurs in Arg1, Arg2, and both, respectively.

- **Word pairs:** We include all word pairs by choosing one word from Arg1 and the other from Arg2.

- **Brown cluster pairs:** Similar to the above features of word pairs, we include all Brown cluster pairs by choosing one word cluster from Arg1 and the other from Arg2.

Besides the above features, the research on English sense classification for non-explicit discourse relations has explored other useful features about polarity, modality, and verb class (Karin et al., 2006). Unfortunately, the shared task on Chinese does not provide relevant resources to obtain those features.

3 Experimentation

We evaluate our system on the Chinese dataset provided in the close track of the CoNLL-2016 Shared Task. All our kernel-based classifiers (e.g., valid EDU pair recognizer, connective recognizer, and linker connecting connectives with EDU pairs) and flat feature-based classifiers (e.g., sense classifiers for either explicit discourse relations or non-explicit discourse relations) are trained using SVMLight toolkit for tree kernel.[4]

Table 1 shows our official performance on the development, test and blind test sets, respectively. From the table, we observe:

- For argument recognition, the performance of Arg2 is much better than that of Arg1 on the development and test datasets. This is similar to the performance trend in English. However, the performance gap between Arg1 and Arg2 recognition is very small on the blind test dataset.

[4]http://disi.unitn.it/moschitti/TK1.0-software/Tree-Kernel.htm

		Dev			Test			Blind test		
		P	R	F1	P	R	F1	P	R	F1
Explicit	Connective	79.22	83.56	81.33	75.00	80.00	77.42	63.07	65.99	64.50
	Arg1	45.45	47.95	46.67	40.62	43.33	41.94	36.57	38.26	37.40
	Arg2	58.44	61.64	60.00	53.12	56.67	54.84	39.05	40.85	39.93
	Arg1 & Arg2	33.77	35.62	34.67	28.12	30.00	29.03	22.79	23.84	23.31
	Overall	35.62	33.77	34.67	27.78	26.04	26.88	21.15	20.14	20.63
Non-Explicit	Connective	-	-	-	-	-	-	-	-	-
	Arg1	65.69	54.32	59.47	62.95	55.67	59.08	54.20	52.36	53.27
	Arg2	72.55	60.00	65.68	69.92	61.82	65.62	55.70	53.81	54.74
	Arg1 & Arg2	55.56	45.95	50.30	52.37	46.31	49.15	42.67	41.22	41.93
	Overall	32.97	39.87	36.09	34.24	38.72	36.34	23.35	24.17	23.75
All	Connective	79.22	83.56	81.33	75.00	80.00	77.42	63.07	65.99	64.50
	Arg1	65.01	56.21	60.29	61.10	56.05	58.46	54.64	53.90	54.27
	Arg2	71.54	61.85	66.34	68.79	63.10	65.83	53.64	52.91	53.27
	Arg1 & Arg2	52.74	45.60	48.91	48.79	44.76	46.69	38.55	38.03	38.29
	Overall	33.77	35.62	34.67	34.07	37.14	35.54	23.31	23.61	23.46

Table 1: Official results (%) of our parser on development, test and blind test sets. Group *Explicit* indicates the performance with respect to explicit discourse relations; group *Non-Explicit* indicates the performance with respect to non-explicit discourse relations, and group *all* indicates the performance with respect to all discourse relations, including both explicit and non-explicit ones.

- With respect to explicit discourse relations, the sense classification works almost perfectly on development data (e.g., almost no performance gap from *Arg1 & Arg2* to *Overall*. It also works well on the test and blind test sets.

- With respect to non-explicit discourse relations, the sense classification works much worse than that of explicit sense classification. The performance gap caused by non-explicit sense classification reaches 14% 18%.

- The overall performance on all discourse relations is dominated by non-explicit ones. This is because larger size of non-explicit discourse relations. For example, the size of non-explicit discourse relations is 3.6 times of that of explicit ones in training data.

- Our system achieves similar results on development set and test set. However, the performance on blind test decreases sharply, probably due to the differences in genres and the bad quality of parse trees.

4 Conclusion

In this paper we have described our submission to the CoNLL-2016 shared task on end-to-end Chinese shallow discourse parsing. Our system is evaluated on the closed track of the CoNLL-2016 shared task and achieves 35.54% and 23.46% in F1-measure on the official test set and blind test set, respectively.

Acknowledgments

The authors would like to thank the two anonymous reviewers for providing helpful suggestions. This research is supported by Project 61472264, 61401295, 61305088 and 61402314 under the National Natural Science Foundation of China.

References

Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: An entity-based approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 141–148.

Kipper Karin, Korhonen Anna, Ryant Neville, and Palmer Martha. 2006. Extending verbnet with novel verb classes. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

Fang Kong, Hwee Tou Ng, and Guodong Zhou. 2014. A constituent-based approach to argument labeling with joint inference in discourse parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 68–77.

Fang Kong, Sheng Li, and Guodong Zhou. 2015. The sonlp-dp system in the conll-2015 shared task. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning - Shared Task*, pages 32–36.

Peifeng Li, Guodong Zhou, Qiaoming Zhu, and Libin Hou. 2012. Employing compositional semantics and discourse consistency in Chinese event extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language*

Processing and Computational Natural Language Learning, pages 1006–1016.

Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014. Building Chinese discourse corpus with connective-driven dependency tree structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 2105–2114.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*.

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: Human Language Technologies*, pages 997–1006.

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2014. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20(2):151–184.

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 International Conference on Language Resources and Evaluation*.

Mei Tu, Yu Zhou, and Chengqing Zong. 2014. Enhancing grammatical cohesion: Generating transitional expressions for SMT. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 850–860.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, and Attapol Rutherford. 2015. The CoNLL-2015 shared task on shallow discourse parsing. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning - Shared Task*, pages 1–16.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Attapol T. Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. 2016. The CoNLL-2016 shared task on multilingual shallow discourse parsing. In *Proceedings of the Twentieth Conference on Computational Natural Language Learning: Shared Task*.

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*, pages 69–77.