

# The SoNLP-DP System in the CoNLL-2015 shared Task

Fang Kong Sheng Li Guodong Zhou

School of Computer Science and Technology, Soochow University, China  
Gangjiang Road 333#, Suzhou, Jiangsu of China 215006

kongfang@suda.edu.cn qcl6355@gmail.com gdzhou@suda.edu.cn

## Abstract

This paper describes the submitted discourse parsing system of the natural language group of Soochow University (SoNLP-DP) to the CoNLL 2015 shared task. Our System classifies discourse relations into explicit and non-explicit relations and uses a pipeline platform to conduct every subtask to form an end-to-end shallow discourse parser in the Penn Discourse Treebank (PDTB). Our system is evaluated on the CoNLL-2015 Shared Task closed track and achieves the 18.51% in F1-measure on the official blind test set.

## 1 Introduction

Discourse parsing determines the internal structure of a text via identifying the discourse relations between its text units and plays an important role in natural language understanding that benefits a wide range of downstream natural language applications, such as coherence modeling (Barzilay and Lapata, 2005; Lin et al., 2011), text summarization (Lin et al., 2012), and statistical machine translation (Meyer and Webber, 2013).

As the largest discourse corpus, the Penn Discourse TreeBank (PDTB) corpus (Prasad et al., 2008) adds a layer of discourse annotations on the top of the Penn TreeBank (PTB) corpus (Marcus et al., 1993) and has been attracting more and more attention recently (Elwell and Baldrige, 2008; Pitler and Nenkova, 2009; Prasad et al., 2010; Ghosh et al., 2011; Kong et al., 2014; Lin et al., 2014). Different from another famous discourse corpus, the Rhetorical Structure Theory(RST) Treebank corpus(Carlson et al., 2001), the PDTB focuses on shallow discourse relations either lexically grounded in explicit discourse connectives or associated with sentential adjacency. This theory-neutral way makes no commitment to

any kind of higher-level discourse structure and can work jointly with high-level topic and functional structuring (Webber et al., 2012) or hierarchical structuring (Asher and Lascarides, 2003).

Although much research work has been conducted for certain subtasks since the release of the PDTB corpus, there is still little work on constructing an end-to-end shallow discourse parser. The CoNLL 2015 shared task (Xue et al., 2015) evaluates end-to-end shallow discourse parsing systems for determining and classifying both explicit and non-explicit discourse relations. A participant system needs to (1)locate all explicit (e.g., "because", "however", "and".) discourse connectives in the text, (2)identify the spans of text that serve as the two arguments for each discourse connective, and (3) predict the sense of the discourse relations (e.g., "Cause", "Condition", "Contrast").

In this paper, we describe the system submission from the NLP group of Soochow university (SoNLP-DP). Our shallow discourse parser consists of multiple components in a pipeline architecture, including a connective classifier, argument labeler, explicit classifier, non-explicit classifier. Our system is evaluated on the CoNLL-2015 Shared Task closed track and achieves the 18.51% in F1-measure on the official blind test set.

The remainder of this paper is organized as follows. Section 2 presents our shallow discourse parsing system. The experimental results are described in Section 3. Section 4 concludes the paper.

## 2 System Architecture

In this section, after a quick overview of our system, we describe the details involved in implementing the end-to-end shallow discourse parser.

### 2.1 System Overview

A typical text consists of sentences glued together in a systematic way to form a coherent discourse.

Referring to the PDTB, shallow discourse parsing focus on shallow discourse relations either lexically grounded in explicit discourse connectives or associated with sentential adjacency. Different from full 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 implicit) taking two arguments (clauses or sentences) with a discourse sense. That is, a complete end-to-end shallow discourse parser includes:

- connective identification, which identifies all connective candidates and labels them as whether they function as discourse connectives or not,
- argument labeling, which identifies the spans of text that serve as the two arguments for each discourse connective,
- explicit sense classification, which predicts the sense of the explicit discourse relations after achieving the connective and its arguments,
- non-explicit sense classification, for all adjacent sentence pairs within each paragraph without explicit discourse relations, which classify the given pair into EntRel, NoRel, or one of the Implicit/AltLex relation senses.

Figure 1 shows the components and the relations among them. Different from the traditional approach (i.e., Lin et al. (2014)), considering the interaction between argument labeler and explicit sense classifier, co-occurrence relation between explicit and non-explicit discourse relations in a text, our system does not employ a complete sequential pipeline framework.

## 2.2 Connective Identification

Our connective identifier works in two steps. First, the connective candidates are extracted from the given text referring to the PDTB. There are 100 types of discourse connectives defined in the PDTB. Then every connective candidate is

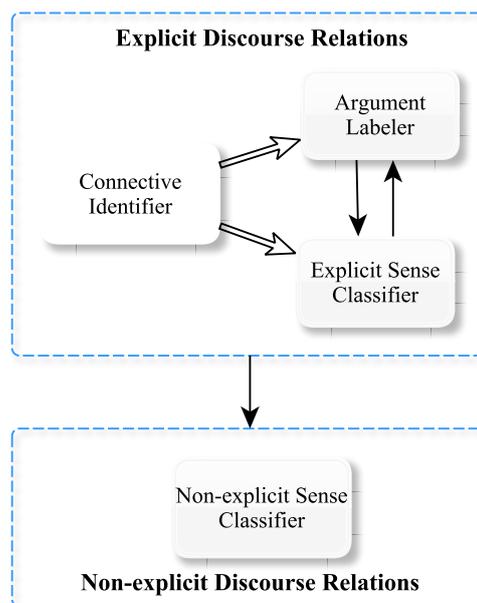


Figure 1: Framework of our end-to-end shallow discourse parser

checked whether it functions as a discourse connective.

Pitler and Nenkova (2009) showed that syntactic features extracted from constituent parse trees are very useful in disambiguating discourse connectives. Followed their work, Lin et al. (2014) found that a connective’s context and part-of-speech (POS) are also helpful. Motivated by their work, we get a set of effective features, includes:

- Lexical: connective itself, POS of the connective, connective with its previous word, connective with its next word, the location of the connective in the sentence, i.e., start, middle and end of the sentence.
- Syntactic: the highest node in the parse tree that covers only the connective words (dominate node), the context of the dominate node <sup>1</sup>, whether the right sibling contains a VP, the path from the parent node of the connective to the root of the parse tree.

Besides, we observed that the syntactic class of the connective<sup>2</sup> and connective modifier (such as

<sup>1</sup>We use POS combination of the parent, left sibling and right sibling of the dominate node to represent the context. When no parent or siblings, it is marked NULL.

<sup>2</sup>All the connectives are classified into four well-defined syntactic classes: subordinating conjunctions, coordinating conjunctions, prepositional phrases and adverbs.

apparently, in large part, etc.) give a very strong indication of its discourse usage. So we introduce both as two additional features.

### 2.3 Argument Labeling

The argument labeler needs to label the Arg1 and Arg2 spans for every connective determined by connective identifier. Following the work of Kong et al. (2014), we employ the constituent-based approach to argument labeling by first extracting the constituents from a parse tree are casted as argument candidates, then determining the role of every constituent as part of Arg1, Arg2, or NULL, and finally, merging all the constituents for Arg1 and Arg2 to obtain the Arg1 and Arg2 text spans respectively.

Specifically, similar to semantic role labeling (SRL), we use a simple algorithm to prune out those constituents that are clearly not arguments to the connective in question. The pruning algorithm works recursively in preprocessing, starting from the target connective node, i.e. the lowest node dominating the connective. First, all the siblings of the connective node are collected as candidates. Then we move on to the parent of the connective node and collect its siblings. This progress goes on until we reach the root of the parse tree.

After extracting the argument candidates, a multi-category classifier is employed to determine the role of every argument candidate (i.e., Arg1, Arg2, or NULL) with features reflecting the properties of the connective, the candidate constituent and relationship between them. Features include,

- Connective related features: connective itself, its syntactic category, its sense class.<sup>3</sup>
- Number of left/right siblings of the connective.
- The context of the constituent. We use POS combination of the constituent, its parent, left sibling and right sibling to represent the context. When there is no parent or siblings, it is marked NULL.
- The path from the parent node of the connective to the node of the constituent.
- The position of the constituent relative to the connective: left, right, or previous.

<sup>3</sup>In training stage, we extract the gold sense class from the annotated corpus. And in testing stage, the sense classification will be employed to get the automatic sense.

### 2.4 Explicit sense classification

After a discourse connective and its two arguments are identified, the sense classifier is proved to decide the sense that the relation conveys.

Although the same connective may carry different semantics under different contexts, only a few connectives are ambiguous (Pitler and Nenkova, 2009). Following the work of Lin et al. (2014), we introduce three features to train a sense classifier: the connective itself, its POS and the previous word of the connective.

Besides, since we observed that various relative positions (i.e., Arg1 precedes Arg2, Arg2 precedes Arg1, Arg2 is embedded within Arg1, or Arg1 is embedded within Arg2) are helpful for sense classification, we includes the relative position as an additional feature.

### 2.5 Non-explicit sense Classification

Referring to the PDTB, the non-explicit relations<sup>4</sup> are annotated for all adjacent sentence pairs within paragraphs. So non-explicit sense classification only considers the sense of every adjacent sentence pair within a paragraph without explicit discourse relations.

Our non-explicit sense classifier includes seven traditional features:

**Verbs:** Following the work of Pitler et al. (2009), we extract the pairs of verbs from the given adjacent sentence pair (i.e., Arg1 and Arg2). Besides that, the number of verb pairs which have the same highest VerbNet verb class (Kipper et al., 2006) is included as a feature. the average length of verb phrases in each argument, and the POS of main verbs are also included.

**Polarity:** This set of features record the number of *positive*, *negated positive*, *negative* and *neutral* words in both arguments and their cross-product. The polarity of every word in arguments is derived from Multi-perspective Question Answering Opinion Corpus(MPQA) (Wilson et al., 2005). Intuitively, polarity features would help recognize Comparison relations.

**Modality:** We include a set of features to record the presence or absence of specific modal words (i.e., can, may, will, shall, must, need) in Arg1 and Arg2, and their cross-product. The intuition

<sup>4</sup>The PDTB provides annotation for Implicit relations, AltLex relations, entity transition (EntRel), and otherwise no relation (NoRel), which are lumped together as Non-Explicit relations.

behind this feature set is that the Contingency relations seem to have more modal words.

**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. Three features are introduced to denote the presence of syntactic productions in Arg1, Arg2 or both. Here, these production rules are extracted from the training data and the rules with frequency less than 5 are ignored.

**Dependency rules:** Similar with Production rules, three features denoting the presence of dependency productions in Arg1, Arg2 or both are also introduced in our system.

**Fisrt/Last and First 3 words:** This set of features include the first and last words of Arg1, the first and last words of Arg2, the pair of the first words of Arg1 and Arg2, the pair of the last words as features, and the first three words of each argument.

**Brown cluster pairs:** We include the Cartesian product of the Brown cluster values of the words in Arg1 and Arg2. In our system, we simply take 100 Brown clusters provided by CoNLL shared task.

Besides, we introduce two features which describe the automatic determined connective list contained by Arg1 and Arg2, respectively, to capture the co-occurrence relationship between non-explicit and explicit discourse relations.

### 3 Experimentation

We train our system on the corpora provided in the CoNLL-2015 Shared Task and evaluate our system on the CoNLL-2015 Shared Task closed track. All our classifiers are trained using the OpenNLP maximum entropy package<sup>5</sup> with the default parameters (i.e. without smoothing and with 100 iterations). We firstly report the official score on the CoNLL-2015 shared task on development, test and blind test sets. Then, the supplementary results provided by the shared task organizes are reported.

In Table 1, we present the official results of our system performances on the CoNLL-2015 development, test and blind test sets, respectively. From the results, we can find that,

- For Connective identification, our system achieved satisfactory results.

<sup>5</sup><http://maxent.sourceforge.net/>

	Development	Test	Blind Test
Arg1&2	43.12	37.01	33.23
Arg1	57.28	52.45	46.28
Arg2	67.72	63.57	61.70
Connective	94.22	94.77	91.62
Sense	17.80	18.38	16.93
Parser	26.32	20.64	18.51

Table 1: the official F1 score of our system.

- For argument labeling, the performance of Arg2 is better than Arg1 and the performance gaps are more than 10% in F1-measure. And the combined results of Arg1 and Arg2 extractor reduced so much in comparison with the performance of Arg1 or Arg2.
- For sense classification, there is a lot of room to improve.
- For the overall parser performance, obviously, a lot of work is needed for end-to-end discourse parsing before practical application.

		Arg1&2	Arg1	Arg2	Sense	Parser
Dev	Exp	34.67	38.67	74.37	20.18	29.78
	nonExp	49.94	62.13	62.37	7.37	23.54
Test	Exp	30.21	34.02	74.48	20.59	25.30
	nonExp	42.38	57.71	54.95	6.77	16.97
Blind	Exp	30.42	36.43	73.04	17.36	22.95
	nonExp	35.87	49.87	51.07	5.24	14.35

Table 2: the supplementary F1 score of our system.

In Table 2, we reported the supplementary results provided by the shared task organizes on the development, test and blind test sets. These additional experiments investigate the performance of our shallow discourse parsing for explicit and non-explicit relations separately. From the results, we can find that the sense classification for both explicit and non-explicit discourse relations are the biggest obstacles to the overall performance of discourse parsing.

### 4 Conclusion

We have presented the SoNLP-DP system from the NLP group of Soochow university that participated in the CoNLL-2015 shared task. Our system is evaluated on the CoNLL-2015 Shared Task closed track and achieves the 18.51% in F1-measure on the official blind test set.

## Acknowledgements

This research is supported by Key project 61333018 and 61331011 under the National Natural Science Foundation of China, Project 6127320 and 61472264 under the National Natural Science Foundation of China.

## References

- Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
- 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*.
- Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of Second SIGdial Workshop on Discourse and Dialogue*.
- Robert Elwell and Jason Baldridge. 2008. Discourse connective argument identification with connective specific rankers. In *Second IEEE International Conference on Semantic Computing*.
- Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi, and Sara Tonelli. 2011. Shallow discourse parsing with conditional random fields. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*.
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending verbnet with novel verb classes. In *Fifth Internal 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*.
- 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*.
- Ziheng Lin, Chang Liu, Hwee Tou Ng, and Min-Yen Kan. 2012. Combining coherence models and machine translation evaluation metrics for summarization evaluation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*.
- 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.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Thomas Meyer and Bonnie Webber. 2013. Implication of discourse connectives in (machine) translation. In *Proceedings of the Workshop on Discourse in Machine Translation*.
- 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*.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the LREC 2008 Conference*.
- Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2010. Exploiting scope for shallow discourse parsing. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*.
- Bonnie Webber, Marcus Egg, and Valia Kordoni. 2012. Discourse structure and language technology. *Natural Language Engineering*, 18(4):437–490, 10.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.
- 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*.