

# A Unified Framework for Discourse Argument Identification via Shallow Semantic Parsing

XU Fan ZHU Qiao Ming\* ZHOU Guo Dong

School of Computer Science and Technology

Soochow University, Suzhou, China 215006

{20104027010, qmzhu, gdzhou}@suda.edu.cn

## ABSTRACT

This paper deals with Discourse Argument Identification (DAI) from both intra-sentence and inter-sentence perspectives. For intra-sentence cases, we approach it via a simplified shallow semantic parsing framework, which recasts the discourse connective as the predicate and its scope into several constituents as the argument of the predicate. Different from state-of-the-art chunking approaches, our parsing approach extends DAI from the chunking level to the parse tree level, where rich syntactic information is available, and focuses on determining whether a constituent, rather than a token, is an argument or not. For inter-sentence cases, we present a lightweight heuristic rule-based solution. Evaluation using Penn Discourse Treebank (PDTB) shows that the current research's parsing approach significantly outperforms the state-of-the-art chunking alternatives.

## TITLE AND ABSTRACT IN CHINESE

### 基于浅层语义分析的篇章论元识别统一框架

本文从句内和句外两种角度处理篇章论元识别问题。针对句内情况，我们采用浅层语义分析框架来处理，将篇章连接词看作谓词，并将谓词的论元映射成一些组块。不同于现有的基于组块方法，我们的语义分析方法将组块层次提升为富于句法信息的句法树层次，同时将组块而不是单词作为处理单元。针对句外情况，我们提出了一种轻量级的规则解决方案。通过宾州篇章树库上的实验，说明我们提出的基于语义分析方法在性能上显著优于现有的基于组块方法。

---

KEYWORDS : Argument Pruning, Discourse Argument Identification, Shallow Semantic Parsing.

KEYWORDS IN CHINESE : 论元过滤, 篇章论元识别, 浅层语义分析.

---

---

\* Corresponding author

## 1 Introduction

Discourse parsing is considered one of the most challenging Natural Language Processing (NLP) tasks. It is essential in many downstream NLP applications, such as statistical machine translation (Meyer, 2011), information retrieval (Huttunen et al., 2011), opinion mining (Somasundaran et al., 2009) and so on.

Generally, Discourse Argument Identification (DAI) involves two sub-tasks: Discourse Connective Identification (DCI) and Argument Scope Identification (ASI). ASI is much more complex than DCI, which has been comprehensively reported in literature with, for example, F-measure of 94.19% on the Penn Discourse Treebank (PDTB) Prasad et al. (2008). Compared with DCI, the performance of ASI is much lower. For example, Ghosh et al. (2012) only got F-measure of 59.39% on exact ArgI identification, on golden tree structures. Most previous studies on DAI focus on token level, such as Ghosh et al. (2012) (2011a) (2011b) and Lin et al. (2010) and suffer from the limitation of focusing on determining whether a token in a discourse simply either belongs to the argument of a connective or not. However, such a strong independence assumption among the tokens may result in poor performance. The tokens should not be independent and sometimes they combine together and play the specific role for the discourse connective.

Accordingly, this paper focuses on PDTB-style exact argument identification, from both intra-sentence and inter-sentence perspectives. For intra-sentence cases, we approach it via a simplified shallow semantic parsing framework, which recasts the discourse connective as the predicate and its scope into several constituents as the argument of the predicate. Different from state-of-the-art chunking approaches, our parsing approach extends DAI from the chunking level to the parse tree level, where rich syntactic information is available, and focuses on determining whether a constituent, rather than a token, is an argument or not. For inter-sentence cases, we present a lightweight heuristic rule-based solution. Evaluation on the PDTB shows that our parsing approach significantly outperforms the afore-mentioned chunking alternatives.

The rest of this paper is organized as follows. Section 2 reviews related work on discourse parsing and on shallow semantic parsing. In Section 3 the PDTB corpus is briefly introduced. Section 4 describes the methodology used for exact argument identification. In Section 5 the results of the research experiment are presented. Finally, some conclusions are drawn.

## 2 Related Work

Related work on PDTB-style discourse parsing and shallow semantic parsing is presented in this section.

PDTB-style discourse parsing consists of two major sub-tasks: Discourse Argument Identification (DAI) and Discourse Relation Identification (DRI). Related work for PDTB-style DAI can be mainly classified into three categories: rule-based approach, Dinesh et al. (2005); Prasad et al., (2010), classification-based method, Wellner et al. (2007); Elwell et al. (2008); Lin et al. (2010) and chunking-based approach, Ghosh et al. (2011a) (2011b) (2012). To be more specific, Dinesh et al. (2005) proposed a tree subtraction method for restricted subordinating connectives. Prasad et al. (2010) provided a set of scope-based filters for argument identification. Wellner et al. (2007) and Elwell et al. (2008) investigated the matching of head-words located in

the argument. However, a potential issue of their work is that no golden head-words were annotated in the PDTB. Lin et al. (2010) proposed a token-level argument node identifier, which determined whether each internal node was an Arg1, Arg2 or Non-argument, and then conducted a tree subtraction algorithm to extract the argument of connectives. Ghosh et al. (2012) which integrated the n-best result of the previous token-level approach (Ghosh et al, 2011a) into their global sentence-level method, significantly improved the method’s DAI performance.

Compared with DAI, explicit and implicit discourse relation identification has been studied more recently, such as Pitler et al. (2009a); Lin et al. (2009); Wang et al. (2010); Zhou et al. (2010); Hong et al. (2012). However, due to inherent difficulties within the implicit discourse relation, its performance is still very low.

Shallow semantic parsing, used to answer ‘the five Ws’ (Who, What, When, Where and Why) questions in a sentence, has been extensively studied in recent years, such as Moschitti (2004) and Li et al. (2010a). Scope learning, a specific shallow semantic parsing problem is also related to DAI. Most existing research on scope learning can be further classified by methodology into rule-based, Chapman et al. (2001), chunking-based, Morante et al. (2009) and shallow semantic parsing-based, Li et al. (2010b) and Zhu et al. (2010).

### 3 Penn Discourse Treebank (PDTB): an Introduction

Currently, PDTB is the largest available discourse corpus. It has annotated 40,600 discourse relations, presented as five relation types: Explicit, Implicit, Alternative Lexicalization (AltLex), Entity-based coherence Relation(EntRel) and No Relation (NoRel). PDTB regards connectives as the discourse predicate, taking two text spans as two arguments, Arg1 and Arg2, which describe the events, facts and/or propositions. Of the two arguments Arg2 is syntactically bound to the connective, while Arg 1 is not. In addition, 3-layered hierarchy, semantic senses have been annotated for Explicit, Implicit and AltLex relations, with 4,16 and 23 kinds of senses for each level, respectively. Due to space limitation, we only give an instance of Explicit relation in this paper. Sentence ‘*In addition, its machines are typically easier to operate, so **customers require less assistance from software.*** (CONTINGENCY: Cause: result).’ (According to PDTB, an Arg1 is indicated by italics, Arg2 indicated in bold, a discourse connective underlined and the sense annotated by parentheses.) is an Explicit instance of where there is an overt connective occurrence between the two arguments.

### 4 Discourse Argument Identification Framework

Similar to Lin et al. (2010), we also separate the DAI into intra-sentence and inter-sentence cases. The entire framework is shown in Figure 1. We run our classifiers in Arg1 position identification and argument identification steps in Figure 1. The performance of DCI is considered reliable, therefore we just integrate AddDiscourse, Pitler et al. (2009b) as the module of connective identification, as shown in Figure 1.

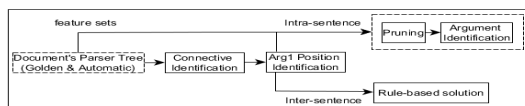


FIGURE 1 – Pipeline framework for discourse argument identification

For Arg1 position identification, Lin et al. (2010) showed that contextual features for connective *C* were useful when identifying Arg1’s position. In addition, we can observe that the first and second next words ( $next_1$  and  $next_2$ ) of *C* also give a strong insight into Arg1’s position. For example, the pronoun ‘that’, as contained within ‘and ensure that’, after the connective ‘and’, sometimes denotes abstract objects located in the previous sentence. Based on this observation, we add 8 new features:  $next_1$ ,  $next_1$  POS,  $next_1+C$ ,  $next_1$  POS+C POS,  $next_2$ ,  $next_2$  POS,  $next_2+C$ ,  $next_2$  POS+C POS. It is hard to decide which feature-set is more effective for Arg1 position identification, even if we use the Hill-climbing (greedy) feature selection technique, Caruana and Freitag (1994), due to the combination of a large number of different features. Therefore, we adopt the Information Gain (IG), which is widely used in text classification, Li et al. (2009), to calculate the efficacy of features and select an approximate optimal feature-set.

After Arg1’s position is identified, we handle the DAI according to intra-sentence and inter-sentence cases methodologies, as follows.

### 4.1 Formulating Intra-sentence DAI as a Simplified Semantic Parsing Problem

Given a parse tree and a predicate, shallow semantic parsing detects and classifies each of the constituents in the sentence into either their corresponding semantic argument (role) for the predicate, or as a non-argument. Similarly, the discourse connective can be taken as the predicate, while its scope can be mapped into several constituents dominated by the connective and thus can be regarded as the argument of the connective. Take this sentence as an example ‘Shorter maturities are considered a sign of rising rates because portfolio managers can capture higher rates sooner.’ The connective ‘because’ has the Arg1 ‘Shorter maturities are considered a sign of rising rates’ and Arg2 ‘portfolio managers can capture higher rates sooner’ the Arg2. As shown in Figure 2 below, the node “IN<sub>9,9</sub>” represents the connective “because” while its Arg1 includes four constituents {NP-SBJ-9<sub>0,1</sub>,VPB<sub>2,2</sub>,VBN<sub>3,3</sub>,S<sub>4,8</sub>}, and its Arg2 includes only one constituent {S<sub>10,16</sub>}. Similar to common shallow semantic parsing, our DAI consist of two pipeline phases: argument pruning and argument identification. Currently, we leave post-processing phase as one of our future works.

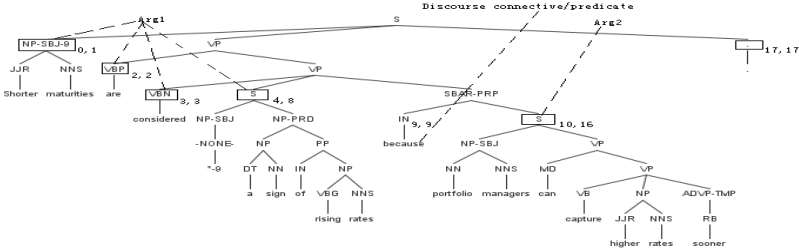


FIGURE 2 – An example of a connective and its argument in a parse tree

**Argument pruning:** Our discourse argument pruning strategy, being similar to that of Xue et al. (2004), which is widely used in common shallow semantic parsing, is detailed as follows.

- (1) Designate the connective as the current node and collect its siblings.
- (2) Reset the current node to its parent and repeat Step (1) until it reaches a threshold *Level* (tree level distance between the current node and the connective).

**Argument identification:** We divide the argument identification into the following two phases. Firstly, a binary classifier is applied to determine whether or not, after pruning, the argument candidates constitute valid arguments. Secondly, a multiclass classifier is adopted to assign a valid argument candidate with a label, e.g. Arg1 or Arg2 or Null.

Most features listed in Table 1 are commonly used in shallow semantic parsing, and most of them are semantic driven (We adopt the head-finding rules described in Collins (1999) and the function type of connection listed in appendix A of Knott (1996).). We categorize the features into three groups as lexical, syntactic and connective-driven features, as shown in Table 1. For the connective-driven features, for example, statistics of connective positions in PDTB tells us that the connective positions are suitable as start, before and back of middle. Therefore, we separate these 3 situations into either lesser or greater than middle, F14, as shown in Table 1.

Features	Remarks	Feature value
<b>Lexical features</b>		
F1	Connective itself	because
F2	Part-of-speech of the connective	IN
F3	The headword and its POS of constituent candidate	sign, NN
F4	The left and right word of the connective	rates, portfolio
F5	The left and right word of the constituent candidate	considered, because
<b>Syntactic features</b>		
F6	The syntactic category of the constituent candidate.	S
F7	The syntactic path from the constituent candidate to the connective.	S<VP>SBAR-PRP>IN
F8	The subcategory of the connective.	SBAR-PRP:IN+S
F9	The phrase type of the connective's parent node.	SBAR-PRP
F10	The subcategory of the constituent candidate.	VP:VBN+S+SBAR-PRP
F11	The phrase type of the constituent candidate's parent node.	VP
F12	The phrase type of the constituent candidate's left and right sibling.	VBN,SBAR-PRP
<b>Connective-driven features</b>		
F13	The position of the constituent candidate with the connective. "left" or "right"	left
F14	The position of the connective in sentence. "lesser than middle" or "greater than middle"	greater than middle
F15	The function type of the connective. "subordinator" or "Coordinator" or "Conj-adverb"	subordinator

TABLE 1 – Features and their instantiations for argument identification within DAI, with “because” as the connective, and S<sub>4,8</sub>, as shown in Figure 2, as the focus constituent.

We don't use features provided by AddDiscourse tool because it was designed for discourse connective and relation identification. In this paper, we formulate the DAI as a shallow semantic

parsing problem, and our main goal is to verify the effectiveness of shallow semantic parsing driven features in DAI. The relationship between the constituents and arguments can be embodied within the shallow semantic parsing framework if we regard connective as predicate.

## 4.2 Rule-based Inter-sentence DAI

According to Prasad et al. (2008)'s statistics, Arg1 in previous, adjacent sentence account for 30.1% in the whole PDTB corpus. Based on this observation, a lightweight heuristic rule-based solution is adopted. Therefore, we assign the preceding adjacent sentence as Arg1, which has already given F1-measures of 76.90% overall in the PDTB, as mentioned in Lin et al. (2010). In addition, we assign the entire sentence excluding the connective as Arg2.

## 5 Experiment

In this section, we describe the experiment settings, together with the experiment results.

### 5.1 Experiment Settings

Similar to Lin et al.(2010)'s evaluation settings, we evaluate our system with GS\_noEP(Gold Standard parsers without Error Propagation), GS\_EP and Auto(Automatic parsers)\_EP settings. All the results use the Johansson and Moschitti (2010)'s exact matching scoring of argument. For the intra-sentence cases, we remove parenthesis and keep subordinate clause for spans to comply with minimality principle on the PDTB, normalize all spans by removing leading or trailing punctuation, and evaluate the system on three main tasks: (1) argument detection, (2)independently classifying phase of known to be discourse arguments into the specific categories, Arg1 and Arg2, and (3) the combined task of detection of the discourse argument and then assigning respective labels (Arg1, Arg2, Null) to them. The Maximum Entropy software package Mallet (<http://mallet.cs.umass.edu/>) is selected as our classifier, and Berkley parser (<http://code.google.com/p/berkeleyparser/downloads/list>) is used to generate the automatic parser tree. For feature selection, we set *IG\_rate* (a threshold value for IG) at a value of 0.5 which is widely used in the common text classification task.

### 5.2 Experiment Results

For the Arg1 position classification phase, we get F-measure of 97.55% using our new added features and feature selection process trained on Section 02-22 and tested on Section 23-24 under GS\_noEP setting. A paired t-test shows that the improvement is significantly superior to Lin's 96.45% with  $p < 0.01$ . The output of final feature sets after using our feature selection are {C; C's position; C POS;  $prev_1$ ;  $prev_1+C$ ;  $prev_1$  POS+C POS;  $prev_2$ ;  $prev_2$  POS;  $next_2$ ;  $next_2$  POS;  $next_2 + C$ ;  $next_2$  POS+C POS}. They show that the new added features along with the feature selection process are useful for determining Arg1's position and they can mitigate the effect of cascaded error propagation.

Experiments on development sets (Section 00-01) show the proper value of *Level* for intra-sentence cases, resulting in an average value of *Level* equal to 3.73. Therefore, we set *Level*=3 and *Level*=4 to check their influence on the parameter *Level* for the argument identification phase. We get F-measure of 86.70% for argument detection when *Level* equals to 4 trained on Section 02-22 and tested on Section 23-24 under GS\_noEP setting, improvements of 1.10% over the *Level* equals to 3. For the heuristic rules in argument pruning, we also tried the pruning strategy

(no Level consideration) as adopted in Xue et al.(2004). However, its performance was about 2.0% lower than our extended pruning strategy. Due to space limitation, we do not give the detail comparison. The result also verifies our assumption that the argument of a connective consists of a constituent in the parser tree, which is always located at a specific level, near the connective.

For the independently classifying phase of Arg1, Arg2 and Both (exact match of both Arg1 and Arg2 simultaneously) for intra-sentence cases, we get Accuracy of 94.15%, 88.72% and 83.28% for Arg1, Arg2 and Both, respectively. With the performance of Level equals to 4 is greater than Level equals to 3, therefore, we conduct the following experiments using a parameter *Level* equals to 4.

Table 2, below, compares the performance for the combined task, between our system and Ghosh’s system. As is shown, the performance of Arg1 exact matching of our system significantly outperforms Ghosh’s system, and the performance of Arg2 with our system slightly outperforms or comparable with theirs. In addition, the performance on automatic syntactic parsing is lower than on the golden parser tree. As expected, some nodes in the automatic parser tree cannot be mapped into corresponding nodes in the golden parser tree, which result in error propagation within the argument pruning and identification phases. Generally, the Precision of Ghosh’s system is higher than our system, while the Recall of their system is lower than ours, which is maybe caused by the features listed in the Table 1 are fine-grained, most of them can capture the relationship between constituent and discourse connective, while features adopted in Ghosh’s system are coarse-grained. Thus, the total F-measure of our system is higher than theirs.

	Our system			Ghosh’s system		
	Arg1	Arg2	Both	Arg1	Arg2	Both
	GS_noEP			(Ghosh et al.,2012)		
Precision(%)	66.28	82.64	58.38	66.10	82.96	-
Recall(%)	59.99	78.24	58.05	53.92	76.28	-
F-measure	<b>62.98*</b>	80.38	58.21	59.39	79.48	-
	GS_EP			(Ghosh et al.,2011b)		
Precision(%)	65.61	83.14	58.44	67.00	82.00	-
Recall(%)	53.36	68.65	50.22	31.00	70.00	-
F-measure	<b>58.85*</b>	75.20	54.02	43.00	76.00	-
	Auto_EP			(Ghosh et al.,2011b)		
Precision(%)	58.45	75.64	56.41	63.00	78.00	-
Recall(%)	40.88	59.12	39.06	28.00	58.00	-
F-measure	<b>48.11*</b>	66.37	46.16	39.00	67.00	-

TABLE 2 – Performance of combined task trained on Section 02-22 and tested on Section 23-24. Performance that is significantly superior to Ghosh’s system ( $p < 0.01$ , using paired t-test for significance) is denoted by \*.

Table 3 illustrates the performance comparison, for combined task, between our system and Lin’s system. As is shown, the performance of Arg1 and Both of our system significantly outperforms Lin’s system under GS\_noEP setting. Furthermore, the performance of Arg1 and Both of our system slightly outperforms Lin’s system under GS\_EP setting. The performance of Arg2 of our system is lower than Lin’s system, which may be caused by the following two reasons: firstly,

Lin et al. (2010) significantly improved the connective identification performance by incorporating their own features and further processing; secondly, the data distribution of intra-sentence and inter-sentence in Section 23 is not coincident with that in Section 23-24. In addition, we can observe that the performance of Arg1 and Both of our system significantly outperforms Lin's system under Auto\_EP setting. This also verifies our framework is robust when facing parser tree error. For example, if the  $S_{10,16}(\text{Arg2})$  of the connective in Figure 2 is incorrectly expanded by the rule  $S_{10,16} \rightarrow \text{NP-SBJ}_{10,11} + \text{MD}_{12,12} + \text{VP}_{13,16}$ , the scope of Arg2 of the connective "because" can still be correctly detected.

	Our system			Lin's system(Lin et al.,2010)		
	Arg1	Arg2	Both	Arg1	Arg2	Both
GS_noEP						
Precision(%)	64.36	83.30	56.16	-	-	-
Recall(%)	57.42	78.66	55.69	-	-	-
F-measure	<b>60.69*</b>	81.00	<b>55.92*</b>	59.15	82.23	53.85
GS_EP						
Precision(%)	62.36	81.15	54.75	-	-	-
Recall(%)	55.04	71.61	51.35	-	-	-
F-measure	58.47	76.08	53.00	57.64	79.80	52.29
Auto_EP						
Precision(%)	62.48	80.86	60.41	-	-	-
Recall(%)	42.36	61.97	40.74	-	-	-
F-measure	<b>50.48*</b>	70.17	<b>48.66*</b>	47.68	70.27	40.37

TABLE 3 – Results of combined task trained on Section 02-21 and tested on Section 23. Performance that is significantly superior to Lin's system ( $p < 0.01$ , using paired t-test for significance) is denoted by \*.

## Conclusions and Future Work

In this paper, we have presented a new approach to PDTB-style discourse argument identification from intra-sentence and inter-sentence perspectives. For the intra-sentence cases, we formulate it as a simplified shallow semantic parsing problem. In particular, we regard the discourse connective as the predicate and map its scope into several constituents, which are deemed as argument of the predicate. For the inter-sentence cases, we present a lightweight heuristic rule-based solution. Evaluation on the PDTB shows the appropriateness of our approach. It also shows that our approach significantly outperforms the state-of-the-art chunking alternatives.

Our future work will be to improve the performance further, through exploring tree kernel-based method, together with more feature engineering.

## Acknowledgments

This research was supported by Projects 60970056, 90920004, 61273320 under the National Natural Science Foundation of China, Project CXZZ11\_0101 under the Program Granted for Scientific Innovation Research of College Graduate of Jiangsu province, Project 2012AA011102 under the National High-Tech Research and Development Plan of China, Project 20093201110006 under the Specialized Research Fund for the Doctoral Program of Higher Education of China, Project BK2011282, 11KJ520003 under the National Natural Science Foundation of Jiangsu province.



## References

- Caruana, R. and Freitag, D. (1994). Greedy attribute selection. In *Proceedings of ML 1994*, pages 28-36.
- Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F. and Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*,34(2001):301-310.
- Collins, M. (1999). Head-driven statistical models for natural language parsing. *Ph.D thesis, USA: University of Pennsylvania*, 1999:1-296.
- Dinesh, N.,Lee, A.,Miltsakaki, E.,Prasad, R. and Joshi, A. (2005). Attribution and the (non-)alignment of syntactic and discourse arguments of connectives. In *Proceedings of Workshop on Frontiers in Corpus Annotation II:Pie in the Sky 2005*, pages 29-36.
- Elwell, R. and Baldridge, J. (2008). Discourse connective argument identification with connective specific rankers. In *Proceedings of ICSC 2008*, pages 198-205.
- Ghosh, S., Johansson, R., Riccardi, G. and Tonelli, S. (2011a). Shallow discourse parsing with conditional random fields. In *Proceedings of IJCNLP 2011*, pages 1071-1079.
- Ghosh, S., Riccardi, G. and Johansson, R. (2012). Global features for shallow discourse parsing. In *Proceedings of SIGDIAL 2012*, pages 150-159.
- Ghosh, S.,Tonelli, S., Riccardi, G. and Johansson, R. (2011b). End-to-end discourse parser evaluation. In *Proceedings of ICSC 2011*, pages 169-172.
- Hong, Y., Zhou, X., Che, T., Yao, J., Zhu, Q. and Zhou, G. (2012). Cross-argument inference for implicit relation extraction. In *Proceedings of CIKM 2012*:accepted.
- Huttunen, S., Vihavainen, A., Etter, P. V. and Yangarber, R. (2011). Relevance prediction in information extraction using discourse and lexical features. In *Proceedings of NCCL 2011*, pages 114-121.
- Johansson, R. and Moschitti, A. (2010). Syntactic and semantic structure for opinion expression detection. In *Proceedings of CoNLL 2010*, pages 67-76.
- Knott, A. (1996). A data-driven methodology for motivating a set of coherence relations. *Ph.D thesis, Scotland: University of Edinburgh*, 1996:1-216.
- Li, S., Xia, R., Zong, C. and Huang, C. (2009). A framework of feature selection methods for text categorization. In *Proceedings of ACL-IJCNLP 2009*, pages 692-700.
- Li, J., Zhou, G., Ng, H. (2010a). Joint Syntactic and Semantic Parsing of Chinese. In *Proceedings of ACL 2010*, pages 1108-1117.
- Li, J., Zhou, G., Wang, H. and Zhu, Q. (2010b). Learning the scope of negation via shallow semantic parsing. In *Proceedings of COLING 2010*, pages 671-679.
- Lin, Z., Kan, M. and Ng H. (2009). Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of EMNLP 2009*, pages 343-351.
- Lin, Z., Ng, H. T. and Kan, M. (2010). A pdtb-styled end-to-end discourse parser. *Technical report TRB8/10, Singapore: National University of Singapore*, 2010:1-15.

- Meyer, T. (2011). Disambiguating temporal-contrastive discourse connectives for machine translation. In *Proceedings of ACL-HLT 2011*, pages 46-51.
- Morante, R. and Daelemans, W. (2009). A metalearning approach to proceeding the scope of negation. In *Proceedings of CoNLL 2009*, pages 21-29.
- Moschitti, A. (2004). A study on convolution kernels for shallow semantic parsing. In *Proceedings of ACL 2004*, pages 335-342.
- Pitler, E., Louis, A. and Nenkova, A. (2009a). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of ACL-IJCNLP 2009*, pages 683-691.
- Pitler, E. and Nenkova, A. (2009b). Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of ACL-IJCNLP 2009*, pages 13-16.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. and Webber, B. (2008). The penn discourse treebank 2.0. In *Proceedings of LREC 2008*, pages 2961-2968.
- Prasad, R., Joshi, A. and Webber, B. (2010). Exploiting scope for shallow discourse parsing. In *Proceedings of LREC 2010*, pages 2076-2083.
- Somasundaran, S., Namata, G., Wiebe, J. and Getoor, L. (2009). Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of EMNLP 2009*, pages 170-179.
- Wang, W., Su, J. and Tan, C. (2010). Kernel based discourse relation recognition with temporal ordering information. In *Proceedings of ACL 2010*, pages 710-719.
- Wellner, B. and Pustejovsky, J. (2007). Automatically identifying the arguments of discourse connectives. In *Proceedings of EMNLP-CoNLL 2007*, pages 92-101.
- Xue, N. and Palmer, M. (2004). Calibrating features for semantic role labeling. In *Proceedings of EMNLP 2004*, pages 88-94.
- Zhou, Z., Xu, Y., Niu, Z., Lan, M., Su, J., Tan, C. (2010). Predicting of discourse connectives for implicit discourse relation recognition. In *Proceedings of COLING 2010*, pages 1507-1514.
- Zhu, Q., Li J., Wang H. and Zhou G. (2010). A unified framework for scope learning via simplified shallow semantic parsing. In *Proceedings of EMNLP 2010*, pages 714-724.