

# Fine-Grained Chinese Discourse Relation Labelling

Huan-Yuan Chen, Wan-Shan Liao, Hen-Hsen Huang, and Hsin-Hsi Chen

Department of Computer Science and Information Engineering, National Taiwan University

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

E-mail: b00902057@ntu.edu.tw, wsliao@nlg.csie.ntu.edu.tw, hhuang@nlg.csie.ntu.edu.tw, hhchen@ntu.edu.tw

## Abstract

This paper explores several aspects together for a fine-grained Chinese discourse analysis. We deal with the issues of ambiguous discourse markers, ambiguous marker linkings, and more than one discourse marker. A universal feature representation is proposed. The pair-once postulation, cross-discourse-unit-first rule and word-pair-marker-first rule select a set of discourse markers from ambiguous linkings. Marker-Sum feature considers total contribution of markers and Marker-Preference feature captures the probability distribution of discourse functions of a representative marker by using preference rule. The HIT Chinese discourse relation treebank (HIT-CDTB) is used to evaluate the proposed models. The 25-way classifier achieves 0.57 micro-averaged F-score.

**Keywords:** Discourse analysis; Discourse relation labelling; HIT Chinese discourse relation treebank.

## 1. Introduction

Discourse relation labelling aims at predicting the most proper discourse relation between two discourse units such as clauses, sentences, and groups of sentences. The labelling task can be done at the intra-sentential and the inter-sentential levels depending on the analysis units. Several schemes have been proposed to define discourse relations to be analyzed. In the PDTB framework (Prasad et al., 2008), three levels of sense hierarchy are utilized. The four classes on the top level are *Temporal*, *Contingency*, *Comparison*, and *Expansion*.

In this paper, we present Chinese discourse analysis on fine-grained discourse relations at intra-sentential level. HIT Chinese discourse relation treebank (Zhang et al., 2014) is used to investigate some specific issues. A discourse marker in Chinese may belong to various lexical categories including conjunction, adverb, noun, preposition, and verb. It may be a single word such as “但” (but) or a word pair such as “雖然...但” (although...but). A discourse marker may have more than one discourse function. Two discourse units may not always be connected with an explicit marker. Besides, a sentence does not always contain only one discourse marker. More than one discourse marker results in an ambiguous-linking issue.

This paper is organized as follows. Section 2 surveys the past works on discourse relation labelling. Section 3 introduces the discourse treebank in this study. Section 4 presents the feature representation and relation labelling algorithm. Section 5 shows the experiments. Section 6 concludes the remarks.

## 2. Related Work

Structure Theory Discourse Treebank (RST-DT) (Carlson et al., 2002) and the Penn Discourse TreeBank (PDTB) (Prasad et al., 2008) are two important resources to facilitate the researches of discourse relation recognition. Because most Chinese discourse corpora follow the

PDTB scheme, this section surveys the related work from this direction.

Pitler and Nenkova (2009) deal with two types of ambiguity of discourse connectives in English: non-discourse vs. discourse usage and unique vs. more discourse functions. They achieve 0.9419 F-score for the first issue and 0.9415 accuracy for the 4-way sense classification of explicit. Hernault et al. (2010, 2011) improve classification accuracy for the infrequent English discourse relation types.

Because the 4-way sense classification accuracy for explicit using just the connective is very high, i.e., 0.9367 (Pitler and Nenkova, 2009), the subsequent works focus on resolution of implicit. Different features are explored. Pitler et al. (2009) use polarity tags, Levin verb classes, length of verb phrases, modality, context, and lexical features. Lin et al. (2009) employ the context of two arguments, word pair information, arguments' internal constituent and dependency parses. Louis et al. (2010) predict implicit discourse relations between adjacent sentences using entity features. Zhou et al. (2010) predict the implicit discourse connective between two arguments with language model. Park and Cardie (2012) optimize feature combinations of the proposed features. Lin et al. (2014) develop an end-to-end discourse parser based on PDTB.

Most of the above works are done for the four top-level classes in PDTB except the paper (Lin et al., 2009). One-vs-the-rest strategy is usually adopted to evaluate the classification performance because the numbers of instances in the classes are unbalanced. Park and Cardie (2012) report the best F-scores for *Temporal* vs. *Rest*, *Contingency* vs. *Rest*, *Comparison* vs. *Rest*, and *Expansion* vs. *Rest* are 0.2657, 0.4982, 0.3132, and 0.7922, respectively. Lin et al. (2009) report the micro-averaged F-score of a fine-grained classification on 11 relation types is 0.40.

Recently, five Chinese discourse corpora (Huang et al., 2013; Zhang et al., 2014; Zhou, Li, et al., 2014; Zhou, Lu, et al., 2014; Li, Feng, et al., 2014) based on PDTB-like

senses have been built. In the CUHK Discourse TreeBank (Zhou, Li, et al., 2014), explicit discourse connectives, their corresponding arguments and senses are annotated. In the HIT-CIR Chinese Discourse Relation Treebank 1.0 (Zhang et al., 2013), coarse-grained and fine-grained relations are labelled at intra- and inter-sentential levels. In NTU discourse corpora (Huang et al., 2013; Huang et al., 2014), 7,601 and 300,000 sentences selected from Chinese part of Clue-Web09 dataset (Yu et al., 2012) are annotated manually and automatically at the intra-sentential level. In the Chinese Discourse Treebank 0.5 released by LDC (Zhou, Lu, et al., 2014), there are approximately 5,500 annotation instances. CDTB 1.0 (Li, Feng, et al., 2014) is composed of 500 news documents. Xue (2005), Zhou and Xue (2012), Li, Feng, et al. (2014) describe the related issues in constructing a Chinese discourse corpus

Chinese connectives are useful, but ambiguous. Zhou et al. (2012) and Li, Carpuat, et al. (2014) use cross-lingual information to disambiguate Chinese discourse connectives. Huang et al. (2014) propose a semi-supervised method to learn probability distribution of discourse functions of connectives, and apply the result to enhance relation labelling.

Most previous Chinese discourse relation labelling was done with coarse-grained senses. Huang and Chen (2011) report F-score of 0.6288 on the 4-way inter-sentential relation classification. Zhou et al. (2012) report the F-score of 0.7481 on 4-way classification at the intra-sentential level. Li, Carpuat et al. (2014) show F-scores for binary classification for *Temporal*, *Contingency*, *Comparison* and *Expansion* are 0.4865, 0.4194, 0.5970 and 0.6920, respectively.

### 3. HIT-CDTB

In HIT-CDTB 1.0 (<http://ir.hit.edu.cn/hit-cdtb/>), discourse relations at intra-sentential, inter-sentential, and passage levels are annotated. It follows the PDTB relation scheme with some modification. The top level consists of 4 discourse relations including *Temporal*, *Contingency*, *Comparison*, and *Expansion*. Each relation is further divided into 2, 9, 5, and 9 subtypes, respectively, shown as follows.

- (1) *Temporal*: asynchronous and precedence.
- (2) *Contingency*: reason first, result first, evidence first, inference first, purpose first, purpose after, sufficient condition first, necessary condition first, and relevance first.
- (3) *Comparison*: forward contrast, reverse contrast, indirect contrast, concession first, and concession after
- (4) *Expansion*: follow, instantiation, example first, example after, exception first, generalization, parallel, compatible selection, and incompatible selection.

Total 525 documents selected from broad-cast news, magazine, newswire, and the web are annotated. Annotators label a relation between two discourse units, and marks explicit or implicit depending on whether there

exists an explicit discourse marker which determines the relation. Table 1 shows the distributions of discourse relations at intra-sentential level. Here, instances of infrequent relation subtypes are removed. The relation distributions are unbalanced. *Expansion* is the majority, and *Temporal* is the minority. Instances of Explicit relation are more than those of Implicit relation.

Relation	#Instances	Explicit	Implicit
Temporal	494	434	60
	5.93%	5.21%	0.72%
Contingency	2,434	1,810	624
	29.23%	21.74%	7.49%
Comparison	1,477	1,301	176
	17.74%	15.63%	2.11%
Expansion	3,921	2,274	1,647
	47.09%	27.31%	19.78%

Table 1: HIT-CDTB at intra-sentential level.

## 4. Chinese Discourse Relation Labelling

### 4.1 Marker Linking Resolution

The discourse relation labelling problem is defined as: given two discourse units,  $ds1$  and  $ds2$ , a relation labeller  $L$  selects a relation  $r$  from a set  $R$  of relations to denote the discourse relation between  $ds1$  and  $ds2$ .  $R$  consists of 4 coarse-grained relations and 25 fine-grained relations.

Determining the correct markers in discourse units is indispensable before using them in relation labelling. The determination is not trivial. (S1) is an example selected from HIT-CDTB. There are three possible word-pair markers, i.e., “雖然...而” (although ... but), “雖然...但” (although ... but), and “是...而非” (is ... but not), and three possible single-word markers, i.e., “而” (ér), “但” (but), and “也” (also). Figure 1 shows the possible marker linkings, where the correct linkings are in bold line, and the other is in dotted line. That is, two word-pair markers and one single-word marker remain.

(S1) [ $ds1$  雖然說八點檔連續劇是商品，而非藝術品],[ $ds2$  但賦予正向的價值觀才算對觀眾負責，也是不斷突破的關鍵]。

([ $ds1$  Although TV serials at 8pm are commodity, but not art], [ $ds2$  the positive values given are responsible for audience, and are also the key to make a breakthrough.])

雖然 ..... 是 ..... ，而非 ..... ，但 ..... ，也 ..... 。

Figure 1: Ambiguous linkings of discourse markers.

We propose three rules as follows to resolve the ambiguous marker linkings:

- (1) pair-once rule: a word can be paired with only one matching word, e.g., “雖然” (although) can be paired with “而” (ér) or “但” (but), but not both.
- (2) cross-discourse-unit-first rule: cross discourse unit > within a discourse unit, e.g., “雖然...但” > “雖然...而”.
- (3) word-pair-marker-first rule: word-pair marker >

single-word marker, e.g., “是...而非”>“而”.

The word “雖然” (although) in  $ds1$  is paired with the word “但” (but) in  $ds2$  by using the second rule. Moreover, the word pair “是...而非” (is ... but not) is more preferred than the single word “而” (ér) by using the third rule. Finally, two word-pair markers, “雖然...但” (although ... but) and “是...而非” (is ... but not), and one single word marker, “也” (also), are selected by using pair-once rule.

## 4.2 More-Than-One-Marker Resolution

We propose two methods, Marker-Sum and Marker-Precedence, to deal with the more-than-one-marker issue. The former aims at representing all the markers appearing in  $ds1$  and  $ds2$ , while the latter tries to select a major marker from  $ds1$  and  $ds2$  by using a precedence rule.

### 4.2.1. Marker-Sum

Each single-word marker is represented as 6 binary features: *Temporal*, *Contingency*, *Comparison*, *Expansion*, *Forward-Linking*, and *Backward-Linking*. Comparatively, only 4 binary features are used to represent a word-pair marker because we do not need to specify forward linking/backward linking.

For each discourse unit, 6 binary features and 4 binary features are allocated for single-word and word-pair marker types, respectively. In a discourse unit, we sum up the feature values of all single-word markers (word-pair markers) in it. We also allocate 4 features to capture the word-pair markers across discourse units. Their values are sum of feature values of the cross discourse-unit markers. Take (S1) as an example.

- $ds1$ : (0,0,0,0,0,0) for single-word marker,
- (0,0,0,1) for word pair “是...而非”,
- $ds2$ : (0,0,0,1,0,1) for single word “也”,
- (0,0,0,0) for word-pair marker,
- $ds1$ - $ds2$ : (0,0,1,0) for word pair “雖然...但”.

### 4.2.2. Marker-Precedence

We select one of the discourse markers appearing in  $ds1$  and  $ds2$  in the precedence of word-pair marker (inter) > word-pair marker (intra) > single-word marker. If there is more than one word-pair marker, we select the first marker from  $ds1$  to  $ds2$ . If there is more than one single-word marker, we select the first marker from  $ds2$  to  $ds1$ .

## 4.3 Relation Labelling Algorithm

We propose the following features to represent a pair of discourse units  $ds1$  and  $ds2$ .

(F1) **Length**. Specify the numbers of words in  $ds1$  and  $ds2$ . The number of **length** features is 2.

(F2) **Punctuation**. Specify which punctuations after  $ds1$  and  $ds2$ . Period, question mark, and exclamation mark are considered. The number of **punctuation** features are 6 (2×3). That is, a binary feature for a punctuation mark in each discourse unit.

(F3) **Shared words**. Specify the number of common

words between  $ds1$  and  $ds2$ . Expansion is based on “同義詞詞林” (<http://ir.hit.edu.cn/>). The number of feature is 1. (F4) **POS**. Specify the frequency of each POS in  $ds1$  and  $ds2$ , respectively. The number of **POS** features is 2×tagging set size.

(F5) **Word unigrams**. Four binary features for each word in the vocabulary are specified: (a) does it occur in  $ds1$ ? (b) does it occur in  $ds2$ ? (c) does it appear at the end of  $ds1$ ? (d) does it appear in the beginning of  $ds2$ ? Total number of **word unigrams** features is 4×vocabulary size.

(F6) **Collocated words**. In training, we compute the PMI of a word pair  $w_i$  and  $w_j$ , where  $w_i$  and  $w_j$  are in  $ds1$  and  $ds2$ , respectively, and select the top 30,000 word pairs of higher PMI. The number of **collocated words** features is 30,000. In testing, we set 1 to a feature when the corresponding collocated words appear.

(F7) **Marker-Sum**. This feature is defined in Section 4.2.1. The number of **Marker-Sum** features is 24, i.e., 10 for  $ds1$ , 10 for  $ds2$ , and 4 for cross discourse unit  $ds1$ - $ds2$ .

(F8) **Marker-Precedence**. This feature is defined in Section 4.2.2. The number of **Marker-Precedence** features is 4.

The relation labelling task is formulated as a multi-way classification. We adopt Scikit-Learn library (Pedregosa et al., 2011) and Logistic Regression as our learning algorithm. When a model is learned, we first run 5-fold cross-validation multiple times on the training set to facilitate a grid search on hyperparameter  $C$ .

## 5. Experiments and Discussions

### 5.1 Coarse-grained Relation Labelling

Table 2 shows the experimental results of 4-way classification. The order of F-scores of relations is: *Expansion* (0.78) > *Comparison* (0.75) > *Contingency* (0.73) > *Temporal* (0.53). The micro-averaged and macro-averaged F-scores are 0.75 and 0.70. Table 3 further examines the effects of individual features. Word unigrams (F5) performs the best F-score. Marker-Sum (F7) ranks the second, in particular, the F-scores for *Contingency*, *Comparison*, and *Expansion* relations are better than those of using (F5). Moreover, Marker-Precedence feature (F8), which copes with the multiple marker problem, ranks the fourth.

Relation	Precision	Recall	F-Score
Temporal	0.46	0.63	0.53
Contingency	0.75	0.71	0.73
Comparison	0.73	0.77	0.75
Expansion	0.80	0.77	0.78
Micro-Avg	0.75	0.74	0.75
Macro-Avg	0.69	0.72	0.70

Table 2: A 4-way classifier on *HIT-CDTB*.

	Tem	Con	Com	Exp	Micro	Macro
F1	0.06	0.29	0.21	0.53	0.38	0.27
F2	0	0.07	0.25	0.62	0.36	0.24
F3	0	0.43	0.24	0.39	0.35	0.27
F4	0.17	0.57	0.74	0.76	0.67	0.56
F5	<b>0.64</b>	<b>0.58</b>	<b>0.51</b>	<b>0.67</b>	<b>0.61</b>	<b>0.60</b>
F6	0.11	0.16	0.15	0.65	0.39	0.27
F7	0.39	<b>0.63</b>	<b>0.61</b>	<b>0.70</b>	<b>0.64</b>	<b>0.58</b>
F8	0.29	0.46	0.35	0.62	0.50	0.43

Table 3: Individual features in a 4-way classifier.

## 5.2 Fine-grained Relation Labelling

We further make a 25-way classification in fine-grained relation labelling. Table 4 lists the number of instances of different relation sub-types and their performance. The micro-averaged F-score is 0.57, less than 0.75 in 4-way classification. The relation subtypes, *purpose first*, *relevance first*, *indirect contrast* and *parallel*, perform above 0.70 F-score. They tend to contain more instances. The subtype, *compatible selection*, achieves 0.63 F-score, although it only contains 74 instances.

The macro-averaged F1-scores of these 25 relation subtypes grouping by *Temporal*, *Contingency*, *Comparison*, and *Expansion* are 0.43, 0.41, 0.35 and 0.45, respectively, with standard deviations 0.15, 0.27, 0.23, and 0.15. The corresponding numbers of subtypes are 2, 9, 5, and 9. We can conclude that the subtypes belonging to *Temporal* and *Expansion* are relatively easier to be identified. Note these are the smallest and the largest groups, respectively.

## 6. Conclusion

In this paper, we propose the *pair-once* rule, *cross-discourse-unit-first* rule and *word-pair-marker-first* rule to select a set of discourse markers from ambiguous linkings, and use the *Marker-Sum* and *Marker-Preference* methods to deal with more than one marker problem. The 25-way classifier achieves 0.57 micro-averaged F-score. Referring to the 11 way implicit relation classifier in English (0.40 micro-averaged F-score), the proposed approach is promising.

## 7. Acknowledgments

This research was partially supported by Ministry of Science and Technology, Taiwan, under grant MOST-102-2221-E-002-103-MY3. We are also thankful to HIT-SCIR for providing us HIT-CIR Chinese Discourse Relation Corpus.

## 8. Bibliographical References

Carlson, L., Marcu, D. and Okurowski, M.E. (2002). RST Discourse Treebank Linguistic Data Consortium, Philadelphia.

Hernault, H., Bollegala, D. and Ishizuka, M. (2010). A Semi-Supervised Approach to Improve Classification of Infrequent Discourse Relations using Feature Vector Extension. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*

relation subtype	Num.	P	R	F1
asynchronous	94	0.33	0.25	0.28
precedence	400	0.53	0.64	0.58
reason first	1,043	0.61	0.51	0.56
result first	250	0.54	0.60	0.57
evidence first	88	0.02	0.01	0.01
inference first	44	0.08	0.07	0.07
<b>purpose first</b>	<b>147</b>	<b>0.78</b>	<b>0.79</b>	<b>0.78</b>
purpose after	229	0.42	0.45	0.43
sufficient condition	74	0.15	0.15	0.15
necessary condition	116	0.36	0.31	0.33
<b>relevance first</b>	<b>443</b>	<b>0.74</b>	<b>0.83</b>	<b>0.78</b>
forward contrast	72	0.38	0.31	0.34
reverse contrast	86	0.09	0.12	0.10
<b>indirect contrast</b>	<b>1,066</b>	<b>0.70</b>	<b>0.71</b>	<b>0.70</b>
concession first	226	0.57	0.61	0.59
concession after	27	0.00	0.00	0.00
follow	1,157	0.55	0.50	0.52
instantiation	632	0.44	0.42	0.43
example first	28	0.20	0.33	0.25
example after	316	0.48	0.48	0.48
exception first	33	0.33	0.43	0.37
generalization	83	0.19	0.42	0.26
<b>parallel</b>	<b>1,468</b>	<b>0.75</b>	<b>0.71</b>	<b>0.73</b>
<b>compatible selection</b>	<b>74</b>	<b>0.52</b>	<b>0.81</b>	<b>0.63</b>
incompatible selection	130	0.29	0.46	0.36
Micro-average	8,326	0.58	0.57	0.57

Table 4: A 25-way classifier on HIT-CDTB.

(EMNLP 2010), pp. 399-409.

- Hernault, F., Bollegala, D., and Ishizuka, M. (2011). Semi-supervised Discourse Relation Classification with Structural Learning. In *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2011)*, Volume Part I, pp. 340-352.
- Huang, H.H. and Chen, H.H. (2011). Chinese Discourse Relation Recognition. In *Proceedings of IJCNLP 2011*, pp. 1442–1446.
- Huang, H.H., Chang, T.W., Chen, H.Y., and Chen, H.H. (2014). Interpretation of Chinese Discourse Connectives for Explicit Discourse Relation Recognition. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, pp. 632–643.
- Huang, H.H., Yu, C.H., Chang, T.W., Lin, C.K., and Chen, H.H. (2013). Analyses of the Association between Discourse Relation and Sentiment Polarity with a Chinese Human-Annotated Corpus. In *Proceedings of 7th Linguistic Annotation Workshop & Interoperability with Discourse*, pp. 70–78.
- Li, Y., Feng, W., Sun, J., Kong, F., and Zhou, G. (2014). Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure. In *Proceedings of the EMNLP 2014*, pp. 2105–2114.
- Li, J.J., Carpuat, M., and Nenkova, A. (2014). Cross-lingual Discourse Relation Analysis: A Corpus Study and A Semi-supervised Classification System. In *Proceedings of COLING 2014*, pp. 577–587.

- Lin, Z., Kan, M.Y. and Ng, H.T. (2009). Recognizing Implicit Discourse Relations in the Penn Discourse Treebank. In *Proceedings of the EMNLP 2009*, pp. 343–351.
- Lin, Z., Ng, H.T. and Kan, M.Y. (2014). A PDTB-styled End-to-end Discourse Parser. *Natural Language Engineering*, 20(2): 151-184.
- Louis, A., Joshi, A., Prasad, R., and Nenkova, A. (2010). Using Entity Features to Classify Implicit Discourse Relations. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2010)*, pp. 59–62.
- Park, J. and Cardie, C. (2012). Improving Implicit Discourse Relation Recognition through Feature Set Optimization. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2012)*, pp. 108–112.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duch-esnay, E. (2011). Scikitlearn: Machine learning in Py-thon. *Journal of Machine Learning Research*, 12, pp. 2825–2830.
- Pitler, E., Louis, A., and Nenkova, A. (2009). Automatic Sense Prediction for Implicit Discourse Relations in Text. In *Proceedings of ACL-IJCNLP 2009*, pp. 683–691.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse Tree-Bank 2.0. In *Proceedings of LREC 2008*, pp. 2961–2968.
- Xue, N. (2005). Annotating Discourse Connectives in the Chinese Treebank. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pp. 84-91.
- Yu, C.H., Tang, Y.J. and Chen, H.H. (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)*, pp. 320-324.
- Zhang, M., Qin, B. and Liu, T. (2014). Chinese Discourse Relation Semantic Taxonomy and Annotation. *Journal of Chinese Information Processing*. 28(2), pp. 28–36.
- Zhou, L., Li, B., Wei, Z., and Wong, K.F. (2014). The CUHK Discourse TreeBank for Chinese: Annotating Explicit Discourse Connectives for the Chinese TreeBank. In *Proceedings of LREC 2014*, pp. 942–949.
- Zhou, L., Gao, W., Li, B., Wei, Z., and Wong, K.F. (2012). Cross-Lingual Identification of Ambiguous Discourse Connectives for Resource-Poor Language. In *Proceedings of COLING 2012*, pp. 1409–1418.
- Zhou, Y., Lu, J., Zhang, J., and Xue, N. (2014). Chinese Discourse Treebank 0.5 LDC2014T21. Web Download. Philadelphia: Linguistic Data Consortium.
- Zhou, Z.M., Xu, Y., Niu, Z.Y., Lan, M., Su, J. and Tan, C.L. (2010). Predicting Discourse Connectives for Implicit Discourse Relation Recognition. In *Proceedings of Coling 2010: Poster Volume*, pp. 1507–1514.
- Zhou, Y. and Xue, N. (2012). PDTB-style Discourse Annotation of Chinese Text. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, pp. 67-77.