

A Cross-lingual Annotation Projection Approach for Relation Detection

Seokhwan Kim[†], Minwoo Jeong[‡], Jonghoon Lee[†], Gary Geunbae Lee[†]

[†]Department of Computer Science and Engineering,
Pohang University of Science and Technology
{megaup | jh21983 | gblee}@postech.ac.kr

[‡]Saarland University
m.jeong@mmci.uni-saarland.de

Abstract

While extensive studies on relation extraction have been conducted in the last decade, statistical systems based on supervised learning are still limited because they require large amounts of training data to achieve high performance. In this paper, we develop a cross-lingual annotation projection method that leverages parallel corpora to bootstrap a relation detector without significant annotation efforts for a resource-poor language. In order to make our method more reliable, we introduce three simple projection noise reduction methods. The merit of our method is demonstrated through a novel Korean relation detection task.

1 Introduction

Relation extraction aims to identify semantic relations of entities in a document. Many relation extraction studies have followed the Relation Detection and Characterization (RDC) task organized by the Automatic Content Extraction project (Doddington et al., 2004) to make multilingual corpora of English, Chinese and Arabic. Although these datasets encourage the development and evaluation of statistical relation extractors for such languages, there would be a scarcity of labeled training samples when learning a new system for another language such as Korean. Since manual annotation of entities and their relations for such *resource-poor languages* is very expensive, we would like to consider instead a weakly-supervised learning technique in

order to learn the relation extractor without significant annotation efforts. To do this, we propose to leverage parallel corpora to project the relation annotation on the source language (e.g. English) to the target (e.g. Korean).

While many supervised machine learning approaches have been successfully applied to the RDC task (Kambhatla, 2004; Zhou et al., 2005; Zelenko et al., 2003; Culotta and Sorensen, 2004; Bunescu and Mooney, 2005; Zhang et al., 2006), few have focused on weakly-supervised relation extraction. For example, (Zhang, 2004) and (Chen et al., 2006) utilized weakly-supervised learning techniques for relation extraction, but they did not consider weak supervision in the context of cross-lingual relation extraction. Our key hypothesis on the use of parallel corpora for learning the relation extraction system is referred to as *cross-lingual annotation projection*. Early studies of cross-lingual annotation projection were accomplished for lexically-based tasks; for example part-of-speech tagging (Yarowsky and Ngai, 2001), named-entity tagging (Yarowsky et al., 2001), and verb classification (Merlo et al., 2002). Recently, there has been increasing interest in applications of annotation projection such as dependency parsing (Hwa et al., 2005), mention detection (Zitouni and Florian, 2008), and semantic role labeling (Pado and Lapata, 2009). However, to the best of our knowledge, no work has reported on the RDC task.

In this paper, we apply a cross-lingual annotation projection approach to binary *relation detection*, a task of identifying the relation between two entities. A simple projection method propagates the relations in source language sentences to

word-aligned target sentences, and a target relation detector can bootstrap from projected annotation. However, this automatic annotation is unreliable because of mis-classification of source text and word alignment errors, so it causes a critical falling-off in annotation projection quality. To alleviate this problem, we present three noise reduction strategies: a heuristic filtering; an alignment correction with dictionary; and an instance selection based on assessment, and combine these to yield a better result.

We provide a quantitative evaluation of our method on a new Korean RDC dataset. In our experiment, we leverage an English-Korean parallel corpus collected from the Web, and demonstrate that the annotation projection approach and noise reduction method are beneficial to build an initial Korean relation detection system. For example, the combined model of three noise reduction methods achieves F1-scores of 36.9% (59.8% precision and 26.7% recall), favorably comparing with the 30.5% shown by the supervised baseline.¹

The remainder of this paper is structured as follows. In Section 2, we describe our cross-lingual annotation projection approach to relation detection task. Then, we present the noise reduction methods in Section 3. Our experiment on the proposed Korean RDC evaluation set is shown in Section 4 and Section 5, and we conclude this paper in Section 6.

2 Cross-lingual Annotation Projection for Relation Detection

The annotation projection from a resource-rich language L_1 to a resource-poor language L_2 is performed by a series of three subtasks: annotation, projection and assessment.

The annotation projection for relation detection can be performed as follows:

- 1) For a given pair of bi-sentences in parallel corpora between a resource-rich language L_1 and a target language L_2 , the relation detection task is carried out for the sentence in L_1 .

¹The dataset and the parallel corpus are available on the author's website, <http://isoft.postech.ac.kr/~megaup/research/resources/>.

- 2) The annotations obtained by analyzing the sentence in L_1 are projected onto the sentence in L_2 based on the word alignment information.
- 3) The projected annotations on the sentence in L_2 are utilized as resources to perform the relation detection task for the language L_2 .

2.1 Annotation

The first step to projecting annotations from L_1 onto L_2 is obtaining annotations for the sentences in L_1 . Since each instance for relation detection is composed of a pair of entity mentions, the information about entity mentions on the given sentences should be identified first. We detect the entities in the L_1 sentences of the parallel corpora. Entity identification generates a number of instances for relation detection by coupling two entities within each sentence. For each instance, the existence of semantic relation between entity mentions is explored, which is called relation detection. We assume that there exist available models or systems for all annotation processes, including not only an entity tagger and a relation detector themselves, but also required preprocessors such as a part-of-speech tagger, base-phrase chunker, and syntax parser for analyzing text in L_1 .

Figure 1 shows an example of annotation projection for relation detection of a bitext in English and Korean. The annotation of the sentence in English shows that "Jan Mullins" and "Computer Recycler Incorporated" are entity mentions of a person and an organization, respectively. Furthermore, the result indicates that the pair of entities has a semantic relationship categorized as "ROLE.Owner" type.

2.2 Projection

In order to project the annotations from the sentences in L_1 onto the sentences in L_2 , we utilize the information of word alignment which plays an important role in statistical machine translation techniques. The word alignment task aims to identify translational relationships among the words in a bitext and produces a bipartite graph with a set of edges between words with translational relationships as shown in Figure 1. In the same manner as the annotation in L_1 , entities are

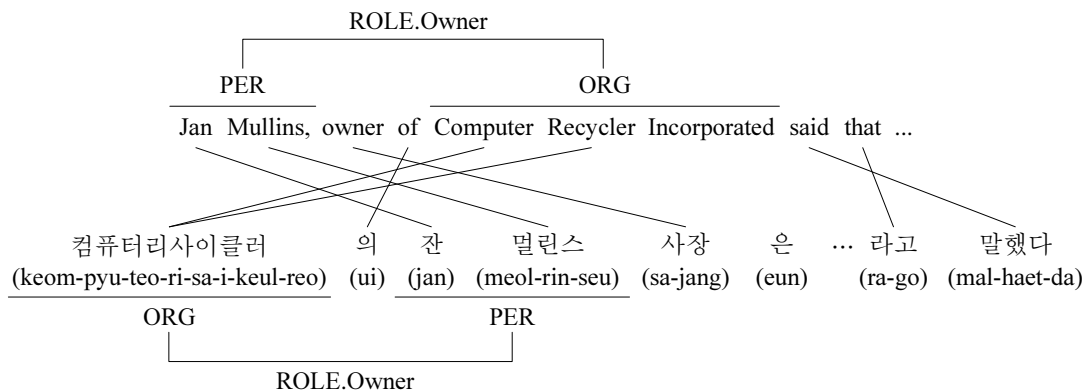


Figure 1: An example of annotation projection for relation detection of a bitext in English and Korean

considered as the first units to be projected. We assume that the words of the sentences in L_2 aligned with a given entity mention in L_1 inherit the information about the original entity in L_1 .

After projecting the annotations of entity mentions, the projections for relational instances follow. A projection is performed on a projected instance in L_2 which is a pair of projected entities by duplicating annotations of the original instance in L_1 .

Figure 1 presents an example of projection of a positive relational instance between “Jan Mullins” and “Computer Recycler Incorporated” in the English sentence onto its translational counterpart sentence in Korean. “Jan meol-rin-seu” and “keom-pyu-teo-ri-sa-i-keul-reo” are labeled as entity mentions with types of a person’s name and an organization’s name respectively. In addition, the instance composed of the two projected entities is annotated as a positive instance, because its original instance on the English sentence also has a semantic relationship.

As the description suggests, the annotation projection approach is highly dependant on the quality of word alignment. However, the results of automatic word alignment may include several noisy or incomplete alignments because of technical difficulties. We present details to tackle the problem by relieving the influence of alignment errors in Section 3.

2.3 Assessment

The most important challenge for annotation projection approaches is how to improve the robust-

ness against the erroneous projections. The noise produced by not only word alignment but also mono-lingual annotations in L_1 accumulates and brings about a drastic decline in the quality of projected annotations.

The simplest policy of utilizing the projected annotations for relation detection in L_2 is to consider that all projected instances are equivalently reliable and to employ entire projections as training instances for the task without any filtering. In contrast with this policy, which is likely to be sub-standard, we propose an alternative policy where the projected instances are assessed and only the instances judged as reliable by the assessment are utilized for the task. Details about the assessment are provided in Section 3.

3 Noise Reduction Strategies

The efforts to reduce noisy projections are considered indispensable parts of the projection-based relation detection method in a resource-poor language. Our noise reduction approach includes the following three strategies: heuristic-based alignment filtering, dictionary-based alignment correction, and assessment-based instance selection.

3.1 Heuristic-based Alignment Filtering

In order to improve the performance of annotation projection approaches, we should break the bottleneck caused by the low quality of automatic word alignment results. As relation detection is carried out for each instance consisting of two entity mentions, the annotation projection for relation detection concerns projecting only entity mentions and

their relational instances. Since this is different from other shallower tasks such as part-of-speech tagging, base phrase chunking, and dependency parsing which should consider projections for all word units, we define and apply some heuristics specialized to projections of entity mentions and relation instances to improve robustness of the method against erroneous alignments, as follows:

- A projection for an entity mention should be based on alignments between contiguous word sequences. If there are one or more gaps in the word sequence in L_2 aligned with an entity mention in the sentence in L_1 , we assume that the corresponding alignments are likely to be erroneous. Thus, the alignments of non-contiguous words are excluded in projection.
- Both an entity mention in L_1 and its projection in L_2 should include at least one base noun phrase. If no base noun phrase occurs in the original entity mention in L_1 , it may suggest some errors in annotation for the sentence in L_1 . The same case for the projected instance raises doubts about alignment errors. The alignments between word sequences without any base noun phrase are filtered out.
- The projected instance in L_2 should satisfy the clausal agreement with the original instance in L_1 . If entities of an instance are located in the same clause (or different clauses), its projected instance should be in the same manner. The instances without clausal agreement are ruled out.

3.2 Dictionary-based Alignment Correction

The errors in word alignment are composed of not only imprecise alignments but also incomplete alignments. If an alignment of an entity among two entities of a relation instance is not provided in the result of the word alignment task, the projection for the corresponding instance is unavailable. Unfortunately, the above-stated alignment filtering heuristics for improving the quality of projections make the annotation loss problems worse by filtering out several alignments likely to be noisy.

In order to solve this problem, a dictionary-based alignment correction strategy is incorporated in our method. The strategy requires a bilingual dictionary for entity mentions. Each entry of the dictionary is a pair of entity mention in L_1 and its translation or transliteration in L_2 . For each entity to be projected from the sentence in L_1 , its counterpart in L_2 is retrieved from the bilingual dictionary. Then, we seek the retrieved entity mention from the sentence in L_2 by finding the longest common subsequence. If a subsequence matched to the retrieved mention is found in the sentence in L_2 , we make a new alignment between it and its original entity on the L_1 sentence.

3.3 Assessment-based Instance Selection

The reliabilities of instances projected via a series of independent modules are different from each other. Thus, we propose an assessment strategy for each projected instance. To evaluate the reliability of a projected instance in L_2 , we use the confidence score of monolingual relation detection for the original counterpart instance in L_1 . The acceptance of a projected instance is determined by whether the score of the instance is larger than a given threshold value θ . Only accepted instances are considered as the results of annotation projection and applied to solve the relation detection task in target language L_2 .

4 Experimental Setup

To demonstrate the effectiveness of our cross-lingual annotation projection approach for relation detection, we performed an experiment on relation detection in Korean text with propagated annotations from English resources.

4.1 Annotation

The first step to evaluate our method was annotating the English sentences in a given parallel corpus. We use an English-Korean parallel corpus crawled from an English-Korean dictionary on the web. The parallel corpus consists of 454,315 bi-sentence pairs in English and Korean². The English sentences in the parallel corpus were prepro-

²The parallel corpus collected and other resources are all available in our website <http://isoft.postech.ac.kr/~megaup/research/resources/>

cessed by the Stanford Parser ³ (Klein and Manning, 2003) which provides a set of analyzed results including part-of-speech tag sequences, a dependency tree, and a constituent parse tree for a sentence.

The annotation for English sentences is divided into two subtasks: entity mention recognition and relation detection. We utilized an off-the-shelf system, Stanford Named Entity Recognizer ⁴ (Finkel et al., 2005) for detecting entity mentions on the English sentences. The total number of English entities detected was 285,566. Each pair of recognized entities within a sentence was considered as an instance for relation detection.

A classification model learned with the training set of the ACE 2003 corpus which consists of 674 documents and 9,683 relation instances was built for relation detection in English. In our implementation, we built a tree kernel-based SVM model using SVM-Light ⁵ (Joachims, 1998) and Tree Kernel Tools ⁶ (Moschitti, 2006). The subtree kernel method (Moschitti, 2006) for shortest path enclosed subtrees (Zhang et al., 2006) was adopted in our model. Our relation detection model achieved 81.2/69.8/75.1 in Precision/Recall/F-measure on the test set of the ACE 2003 corpus, which consists of 97 documents and 1,386 relation instances.

The annotation of relations was performed by determining the existence of semantic relations for all 115,452 instances with the trained model for relation detection. The annotation detected 22,162 instances as positive which have semantic relations.

4.2 Projection

The labels about entities and relations in the English sentences of the parallel corpora were propagated into the corresponding sentences in Korean. The Korean sentences were preprocessed by our part-of-speech tagger ⁷ (Lee et al., 2002) and a dependency parser implemented by MSTParser with

³<http://nlp.stanford.edu/software/lex-parser.shtml>

⁴<http://nlp.stanford.edu/software/CRF-NER.shtml>

⁵<http://svmlight.joachims.org/>

⁶<http://disi.unitn.it/~moschitt/Tree-Kernel.htm>

⁷<http://isoft.postech.ac.kr/~megaup/research/postag/>

| Filter | Without assessing | With assessing |
|--------------|-------------------|----------------|
| none | 97,239 | 39,203 |
| + heuristics | 31,652 | 12,775 |
| + dictionary | 39,891 | 17,381 |

Table 1: Numbers of projected instances

a model trained on the Sejong corpus (Kim, 2006).

The annotation projections were performed on the bi-sentences of the parallel corpus followed by descriptions mentioned in Section 2.2. The bi-sentences were processed by the GIZA++ software (Och and Ney, 2003) in the standard configuration in both English-Korean and Korean-English directions. The bi-directional alignments were joined by the grow-diag-final algorithm, which is widely used in bilingual phrase extraction (Koehn et al., 2003) for statistical machine translation. This system achieved 65.1/41.6/50.8 in Precision/Recall/F-measure in our evaluation of 201 randomly sampled English-Korean bi-sentences with manually annotated alignments.

The number of projected instances varied with the applied strategies for reducing noise as shown in Table 1. Many projected instances were filtered out by heuristics, and only 32.6% of the instances were left. However, several instances were rescued by dictionary-based alignment correction and the number of projected instances increased from 31,652 to 39,891. For all cases of noise reduction strategies, we performed the assessment-based instance selection with a threshold value θ of 0.7, which was determined empirically through the grid search method. About 40% of the projected instances were accepted by instance selection.

4.3 Evaluation

In order to evaluate our proposed method, we prepared a dataset for the Korean RDC task. The dataset was built by annotating the information about entities and relations in 100 news documents in Korean. The annotations were performed by two annotators following the guidelines for the ACE corpus processed by LDC. Our Korean RDC corpus consists of 835 sentences, 3,331 entity mentions, and 8,354 relation instances. The sen-

| Model | w/o assessing | | | with assessing | | |
|------------------------|---------------|------|------|----------------|------|------|
| | P | R | F | P | R | F |
| Baseline | 60.5 | 20.4 | 30.5 | - | - | - |
| Non-filtered | 22.5 | 6.5 | 10.0 | 29.1 | 13.2 | 18.2 |
| Heuristic | 51.4 | 15.5 | 23.8 | 56.1 | 22.9 | 32.5 |
| Heuristic + Dictionary | 55.3 | 19.4 | 28.7 | 59.8 | 26.7 | 36.9 |

Table 2: Experimental Results

tences of the corpus were preprocessed by equivalent systems used for analyzing Korean sentences for projection. We randomly divided the dataset into two subsets with the same number of instances for use as a training set to build the baseline system and for evaluation.

For evaluating our approach, training instance sets to learn models were prepared for relation detection in Korean. The instances of the training set (half of the manually built Korean RDC corpufs) were used to train the baseline model. All other sets of instances include these baseline instances and additional instances propagated by the annotation projection approach. The training sets with projected instances are categorized into three groups by the level of applied strategies for noise reduction. While the first set included all projections without any noise reduction strategies, the second included only the instances accepted by the heuristics. The last set consisted of the results of a series of heuristic-based filtering and dictionary-based correction. For each training set with projected instances, an additional set was derived by performing assessment-based instance selection.

We built the relation detection models for all seven training sets (a baseline set, three projected sets without assessing, and three projected sets with assessing). Our implementations are based on the SVM-Light and Tree Kernel Tools described in the former subsection. The shortest path dependency kernel (Bunescu and Mooney, 2005) implemented by the subtree kernel method (Moschitti, 2006) was adopted to learn all models.

The performance for each model was evaluated with the predictions of the model on the test set, which was the other half of Korean RDC corpus.

We measured the performances of the models on true entity mentions with true chaining of coreference. Precision, Recall and F-measure were adopted for our evaluation.

5 Experimental Results

Table 2 compares the performances of the different models which are distinguished by the applied strategies for noise reduction. It shows that:

- The model with non-filtered projections achieves extremely poor performance due to a large number of erroneous instances. This indicates that the efforts for reducing noise are urgently needed.
- The heuristic-based alignment filtering helps to improve the performance. However, it is much worse than the baseline performance because of a falling-off in recall.
- The dictionary-based correction to our projections increased both precision and recall compared with the former models with projected instances. Nevertheless, it still fails to achieve performance improvement over the baseline model.
- For all models with projection, the assessment-based instance selection boosts the performances significantly. This means that this selection strategy is crucial in improving the performance of the models by excluding unreliable instances with low confidence.
- The model with heuristics and assessments finally achieves better performance than the baseline model. This suggests that the projected instances have a beneficial influence

on the relation detection task when at least these two strategies are adopted for reducing noises.

- The final model incorporating all proposed noise reduction strategies outperforms the baseline model by 6 in F-measure. This is due to largely increased recall by absorbing more useful features from the well-refined set of projected instances.

The experimental results show that our proposed techniques effectively improve the performance of relation detection in the resource-poor Korean language with a set of annotations projected from the resource-rich English language.

6 Conclusion

This paper presented a novel cross-lingual annotation projection method for relation extraction in a resource-poor language. We proposed methods of propagating annotations from a resource-rich language to a target language via parallel corpora. In order to relieve the bad influence of noisy projections, we focused on the strategies for reducing the noise generated during the projection. We applied our methods to the relation detection task in Korean. Experimental results show that the projected instances from an English-Korean parallel corpus help to improve the performance of the task when our noise reduction strategies are adopted.

We would like to introduce our method to the other subtask of relation extraction, which is relation categorization. While relation detection is a binary classification problem, relation categorization can be solved by a classifier for multiple classes. Since the fundamental approaches of the two tasks are similar, we expect that our projection-based relation detection methods can be easily adapted to the relation categorization task.

For this further work, we are concerned about the problem of low performance for Korean, which was below 40 for relation detection. The relation categorization performance is mostly lower than detection because of the larger number of classes to be classified, so the performance of projection-based approaches has to be improved

in order to apply them. An experimental result of this work shows that the most important factor in improving the performance is how to select the reliable instances from a large number of projections. We plan to develop more elaborate strategies for instance selection to improve the projection performance for relation extraction.

Acknowledgement

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2010-C1090-1031-0009).

References

- Bunescu, Razvan C. and Raymond J. Mooney. 2005. A shortest path dependency kernel for relation extraction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, page 724731.
- Chen, Jinxiu, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. 2006. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 129–136, Sydney, Australia. Association for Computational Linguistics.
- Culotta, Aron and Jaffrey Sorensen. 2004. Dependency tree kernels for relation extraction. In *Proceedings of ACL*, volume 4.
- Doddington, George, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) programtasks, data, and evaluation. In *Proceedings of LREC*, volume 4, page 837840.
- Finkel, Jenny R., Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, volume 43, page 363.
- Hwa, Rebecca, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural Language Engineering*, 11(03):311–325.

- Joachims, Thorsten. 1998. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the European Conference on Machine Learning*, pages 137–142.
- Kambhatla, Nanda. 2004. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 22, Barcelona, Spain. Association for Computational Linguistics.
- Kim, Hansaem. 2006. Korean national corpus in the 21st century sejong project. In *Proceedings of the 13th NIJL International Symposium*, page 4954.
- Klein, Dan and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 423–430, Sapporo, Japan. Association for Computational Linguistics.
- Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, volume 1, pages 48–54.
- Lee, Gary Geunbae, Jeongwon Cha, and Jong-Hyeok Lee. 2002. Syllable pattern-based unknown morpheme segmentation and estimation for hybrid part-of-speech tagging of korean. *Computational Linguistics*, 28(1):53–70.
- Merlo, Paola, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. 2002. A multilingual paradigm for automatic verb classification. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 207–214, Philadelphia, Pennsylvania. Association for Computational Linguistics.
- Moschitti, Alessandro. 2006. Making tree kernels practical for natural language learning. In *Proceedings of EACL06*.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, March.
- Pado, Sebastian and Mirella Lapata. 2009. Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, 36(1):307340.
- Yarowsky, David and Grace Ngai. 2001. Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001*, pages 1–8, Pittsburgh, Pennsylvania. Association for Computational Linguistics.
- Yarowsky, David, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on Human language technology research*, pages 1–8, San Diego. Association for Computational Linguistics.
- Zelenko, Dmitry, Chinatsu Aone, and Anthony Richardella. 2003. Kernel methods for relation extraction. *J. Mach. Learn. Res.*, 3:1083–1106.
- Zhang, Min, Jie Zhang, Jian Su, and Guodong Zhou. 2006. A composite kernel to extract relations between entities with both flat and structured features. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 825–832, Sydney, Australia. Association for Computational Linguistics.
- Zhang, Zhu. 2004. Weakly-supervised relation classification for information extraction. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 581–588, Washington, D.C., USA. ACM.
- Zhou, Guodong, Jian Su, Jie Zhang, and Min Zhang. 2005. Exploring various knowledge in relation extraction. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, page 434.
- Zitouni, Imed and Radu Florian. 2008. Mention detection crossing the language barrier. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 600–609, Honolulu, Hawaii. Association for Computational Linguistics.