Semantic Role Labeling of Nominalized Predicates in Chinese

Nianwen Xue
Center for Research in Spoken Language
University of Colorado
Boulder, CO, 80309
`Nianwen.Xue@colorado.edu`

Abstract

Recent work on semantic role labeling (SRL) has focused almost exclusively on the analysis of the predicate-argument structure of verbs, largely due to the lack of human-annotated resources for other types of predicates that can serve as training and test data for the semantic role labeling systems. However, it is well-known that verbs are not the only type of predicates that can take arguments. Most notably, nouns that are nominalized forms of verbs and relational nouns generally are also considered to have their own predicate-argument structure. In this paper we report results of SRL experiments on nominalized predicates in Chinese, using a newly completed corpus, the Chinese Nombank. We also discuss the impact of using publicly available manually annotated verb data to improve the SRL accuracy of nouns, exploiting a widely-held assumption that verbs and their nominalizations share the same predicate-argument structure. Finally, we discuss the results of applying reranking techniques to improve SRL accuracy for nominalized predicates, which showed insignificant improvement.

1   Introduction

Detecting and classifying the arguments of predicates has been an active area of research in recent years, driven by the availability of large-scale semantically annotated corpora such as the FrameNet (Baker et al., 1998) and the Propbank (Palmer et al., 2005). It is generally formulated as a semantic role labeling (SRL) task, where each argument of the predicate is assigned a label that represents the semantic role it plays with regard to its predicate (Gildea and Jurafsky, 2002; Hacioglu et al., 2003; Pradhan et al., 2004b; Xue and Palmer, 2004; Toutanova et al., 2005; Koomen et al., 2005). It has been the shared task for the CoNLL competition for two consecutive years (Carreras and Màrquez, 2004b; Carreras and Màrquez, 2005). This line of research has also expanded from English to other languages (Sun and Jurafsky, 2004; Xue and Palmer, 2005). So far, however, most of the research efforts have focused on analyzing the predicate-argument structure of verbs, largely due to absence of annotated data for other predicate types. In this paper, we report SRL experiments performed on nominalized predicates in Chinese, taking advantage of a newly completed corpus, the Chinese Nombank (Xue, 2006), which we describe in greater detail in Section 2. The rest of the paper is organized as follows. Section 3 describes the architecture of our system as well as the features we used in our experiments. In Section 4 we describe the experimental setups and report our experimental results. We first present experiments that use hand-crafted parses as input, providing a measurement of how well the Nombank annotation can be bootstrapped from the syntactic structure in the treebank. We then describe a more realistic experimental setup in which an automatic parser is first used to parse unsegmented raw

text and its output is then fed into our SRL system. We also discuss whether verb data can be used to improve the SRL accuracy of nominalized predicates. Finally we describe a preliminary experiment that uses reranking techniques to improve the SRL accuracy on hand-crafted parses. Section 5 attempts to put our results in perspective in the context of related work. Section 6 concludes our paper.

2 The Chinese Nombank

The Chinese Nombank extends the general annotation framework of the English Proposition Bank (Palmer et al., 2005) and the English Nombank (Meyers et al., 2004) to the annotation of nominalized predicates in Chinese. Like the English Nombank project, the Chinese Nombank adds a layer of semantic annotation to the Chinese Tree-Bank (CTB), a syntactically annotated corpus of 500 thousand words. The Chinese Nombank annotates two types of elements that are associated with the nominalized predicate: argument-like elements that are expected of this predicate, and adjunct-like elements that modify this predicate. Arguments are assigned numbered labels (prefixed by *ARG*, e.g., ARG0...ARGn) while adjuncts receive a functional tag (e.g., TMP for temporal, LOC for locative, MNR for manner) prefixed by *ARGM*. A predicate generally has no more than six numbered arguments and the complete list of functional tags for adjuncts and their descriptions can be found in the annotation guidelines of this project.

The Chinese Nombank also adds a coarse-grained sense tag to the predicate. The senses of a predicate, formally called *framesets*, are motivated by the argument structure of this predicate and are thus an integral part of the predicate-argument structure annotation. Sense disambiguation is performed only when different senses of a predicate require different sets of arguments. These senses are the same senses defined for the corresponding verbs in the Chinese Proposition Bank, but typically only a subset of the verb senses are realized in their nominalized forms. The example in 1 illustrates the Chinese Nombank annotations, which are the labels in bold in the parse tree. Take 发展("development") as an example, **f1** is the frameset identifier. Of the four expected arguments for this frameset, **ARG0** the cause or agent,
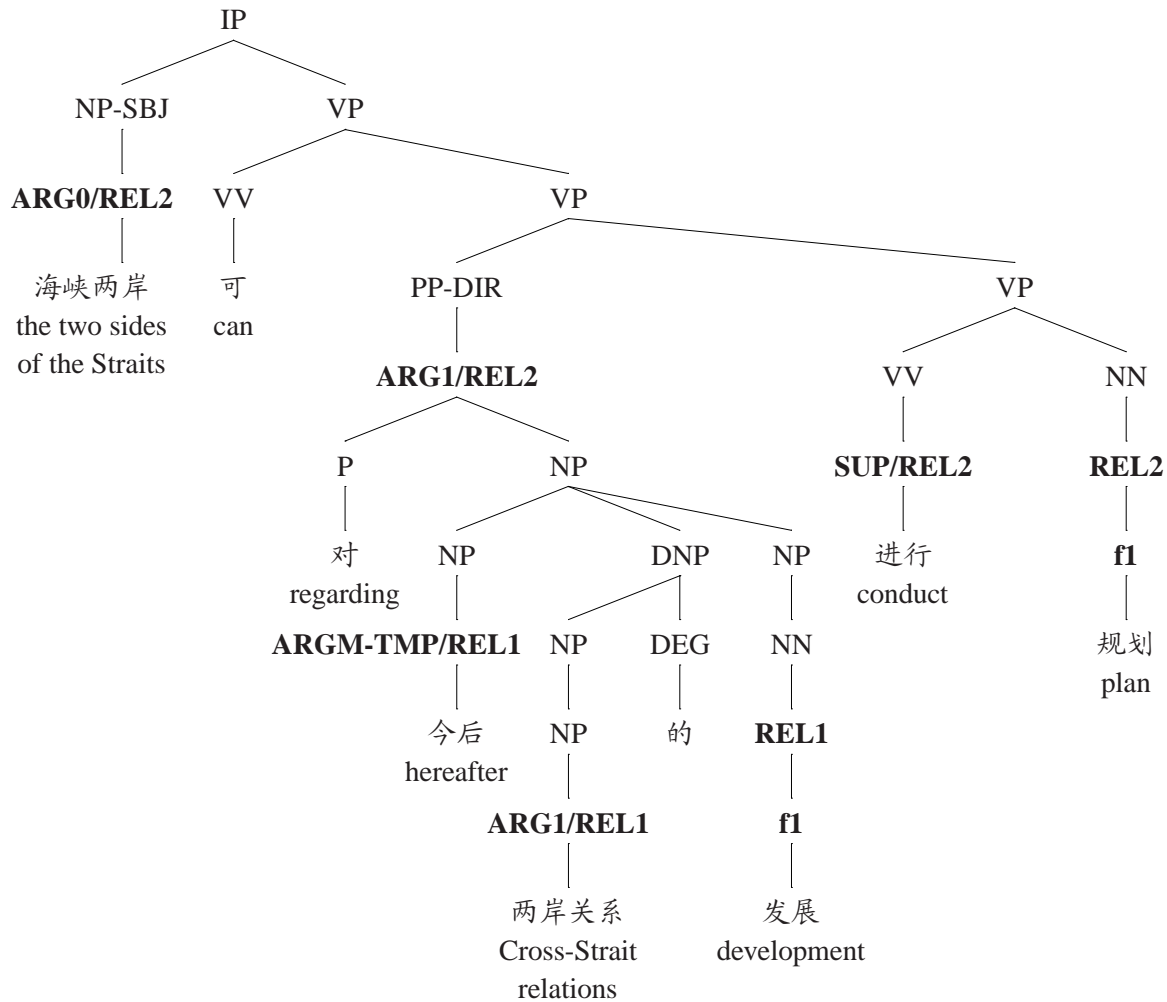
ARG1 the theme, **ARG2** the initial state and **ARG3** the end state or goal, only ARG1 is realized and it is 两岸关系("cross-Strait relations"). The predicate also has a modifier labeled **ARGM-TMP**, 今后("hereafter").

Typically the arguments and adjuncts of a nominalized predicate are realized inside the noun phrase headed by the nominalized predicate, as is the case for 发展("development") in Example 1. A main exception is when the noun phrase headed by the nominalized predicate is an object of a support verb, in which case the arguments of this predicate can occur outside the noun phrase. This is illustrated by 规划("planning") in Example 1, where the noun phrase of which it is the head is the object of a support verb 进行("conduct"), which has little meaning of its own. Both arguments of this predicate, 海峡两岸("the two sides of the Taiwan Strait") and 今后两岸关系的发展("the development of the cross-Strait relations"), are realized outside the noun phrase. There are also a few other general tendencies about the arguments of nominalized predicates that are worth pointing out. The distribution of their arguments is much less predictable than verbs whose arguments typically occupy prominent syntactic positions like the subject and object. There also tend to be fewer arguments that are actually realized for nominalized predicates. Nominalized predicates also tend to take fewer types of adjuncts (**ARGM**s) than their verbal counterpart and they also tend to be less polysemous, having only a subset of the senses of their verb counterpart.

The goal of the semantic role labeling task described in this paper is to identify the arguments and adjuncts of nominalized predicates and assign appropriate semantic role labels to them. For the purposes of our experiments, the sense information of the predicates are ignored and left for future research.

3 System description

The predominant approach to the semantic role labeling task is to formulate it as a classification problem that can be solved with machine-learning techniques. Argument detection is generally formulated as a binary classification task that separates constituents that are arguments or adjuncts to a pred-

432

IP
NP-SBJ　　VP
ARG0/REL2　VV　　VP
　　　　　　　　PP-DIR　　　　VP
　　　　　　　　ARG1/REL2　　VV　　NN
　　　　　　　　P　　NP　　**SUP/REL2**　**REL2**
海峡两岸　可　　对　NP　　DNP　NP　进行　**f1**
the two sides　can　regarding　**ARGM-TMP/REL1**　NP　DEG　NN　conduct　规划
of the Straits　　　　　　今后　NP　的　**REL1**　plan
　　　　　　　　　　　hereafter　**ARG1/REL1**　**f1**
　　　　　　　　　　　　两岸关系　发展
　　　　　　　　　　　　Cross-Strait　development
　　　　　　　　　　　　relations

The two sides of the Taiwan Straits can plan the development of the cross-Strait relations hereafter.

Table 1: A nominalized predicate annotated with semantic roles

icate from those that are not related to the predicate in question. Argument classification, which classifies the constituents into a category that corresponds to one of the argument or adjunct labels is a natural multi-category classification problem. Many classification techniques, SVM (Pradhan et al., 2004b), perceptrons (Carreras and Màrquez, 2004a), Maximum Entropy (Xue and Palmer, 2004), etc. have been successfully used to solve SRL problems. For our purposes here, we use a Maximum Entropy classifier with a tunable Gaussian prior in the Mallet Toolkit[1]. The Maximum Entropy classifier does multi-category classification and thus can be straightforwardly applied to the problem here. The classifier can be tuned to minimize overfitting by adjusting the Gaussian prior.

3.1 A three-stage architecture

Like verbal predicates, the arguments and adjuncts of a nominalized predicate are related to the predicate itself in linguistically well-understood structural configurations. As we pointed out in Section 2, most of the arguments for nominalized predicates are inside the NP headed by the predicate unless the NP is the object of a support verb, in which case its arguments can occur outside the NP. Typically the subject of the support verb is also an argument of the nominalized predicate, as illustrated in Example 1.

---

[1]http://mallet.cs.umass.edu

433

The majority of the constituents are not related to the predicate in question, especially since the sentences in the treebank tend to be very long. This is clearly a lingustic observation that can be exploited for the purpose of argument detection. There are two common approaches to argument detection in the SRL literature. One is to apply a binary classifier directly to all the constituents in the parse tree to separate the arguments from non-arguments, and let the machine learning algorithm do the work. This can be done with high accuracy when the machine-learning algorithm is powerful and is provided with appropriate features (Hacioglu et al., 2003; Pradhan et al., 2004b). The alternative approach is to combine heuristic and machine-learning approaches (Xue and Palmer, 2004). Some negative samples are first filtered out with heuristics that exploit the syntactic structures represented in a parse tree before a binary classifier is applied to further separate the positive samples from the negative samples. It turns out the heuristics that are first proposed in Xue and Palmer (2004) to prune out non-arguments for verbal predicates can be easily adapted to detect arguments for the nominalized predicates as well, so in our experiments we adopt the latter approach. The algorithm starts from the predicate that anchors the annotation, and first collects all the sisters of this predicate. It then iteratively moves one level up to the parent of the current node to collect its sisters till it reaches the appropriate top-level node. At each level, the system has a procedure to determine whether that level is a coordination structure or a modification structure. The system only considers a constituent to be a potential candidate if it is an adjunct to the current node. Punctuation marks at all levels are skipped. After this initial procedure, a binary classifier is applied to distinguish the positive samples from the negative samples. A lower threshold is used for positive samples than negative samples to maximize the recall so that we can pass along as many positive samples as possible to the next stage, which is the multi-category classification.

3.2 Features

SRL differs from low-level NLP tasks such as POS tagging in that it has a fairly large feature space and as a result linguistic knowledge is crucial in designing effective features for this task. A wide range of features have been shown to be useful in previous work on semantic role labeling for verbal predicates (Gildea and Jurafsky, 2002; Pradhan et al., 2004b; Xue and Palmer, 2004) and our experiments show most of them are also effective for SRL of nominalized predicates. The features for our multicategory classifier are listed below:

- *Predicate*: The nominalized predicate itself.

- *Position*: The position is defined in relation to the predicate and the values are *before* and *after*. Since most of the arguments for nominalized predicates in Chinese are before the predicates, this feature is not as effective as when it is used for verbal predicates.

- *path*: The path between the constituent being classified and the predicate.

- *path + dominating verb*. The path feature combined with the dominating verb. This feature is only invoked when there is an intervening dominating verb between the constituent being classified and the predicate. It is used to capture the observation that only a closed set of verbs can be support verbs for nominalized predicates and they are good indicators of whether or not the constituent is an argument of this predicate and the semantic role of the argument.

- *Head word and its part of speech*: The head word and its part-of-speech have proved to be a good indicator of the semantic role label of a constituent for verbal predicates in previous work. It proves to be a good feature for nominal predicates as well.

- *Phrase type*: The syntactic category of the constituent being classified.

- *First and last word of the constituent being classified*

- *sisterhood with predicate*: A binary feature that indicates whether the constituent being classified is a sister to the nominalized predicate.

- *Combination features*: predicate-head word combination, predicate-phrase type combination.

- *class features*. Features that replace the predicate with its class. The class features are induced from frame files through a procedure first introduced in (Xue and Palmer, 2005).

Not all the features used for multicategory classification are equally effective for binary classification, which only determines whether or not a constituent is an argument or adjunct to the nominalized predicate. Therefore, the features for the binary classifier are a subset of the features used for multicategory classification. These are path, path plus dominating verb, head word and its part-of-speech and sisterhood.

4 Experiments

4.1 Data

Our system is trained and tested on a pre-release version of the Chinese Nombank. This version of the Chinese Nombank consists of standoff annotation on the first 760 articles (`chtb_001.fid` to `chtb_931.fid`) of the Penn Chinese Treebank[2]. This chunk of data has 250K words and 10,364 sentences. It has 1,227 nominalized predicate types and 10,497 nominalized predicate instances. In comparison, there are 4,854 verb predicate types and 37,183 verb predicate instances in the same chunk of data. By instance, the size of the Nombank is between a quarter and one third of the Chinese Proposition Bank. Following the convention of the semantic role labeling experiments in previous work, we divide the training and test data by the number of articles, not by the predicate instances. This pretty much guarantees that there will be unseen predicates in the test data. For all our experiments, 688 files are used as training data and the other 72 files (`chtb_001.fid` to `chtb_040.fid` and `chtb_900.fid` to `chtb_931.fid`) are held out as test data. The test data is selected from the double-annotated files in the Chinese Treebank and the complete list of double-annotated files can be found in the documentation for the Chinese Treebank 5.1. Our parser is trained and tested with the same data partition as our semantic role labeling system.

4.2 Semantic role tagging with hand-crafted parses

In this section we present experimental results using Gold Standard parses in the Chinese Treebank as input. To be used in real-world natural language applications, a semantic role tagger has to use automatically produced constituent boundaries either from a parser or by some other means, but experiments with Gold Standard input will help us evaluate how much of a challenge it is to map a syntactic representation to a semantic representation, which may very well vary from language to language. There are two experimental setups. In the first experiment, we assume that the constituents that are arguments or adjuncts are known. We only need to assign the correct argument or adjunct labels. In the second experiment, we assume that all the constituents in a parse tree are possible arguments. The system first filters out consituents that are highly unlikely to be an argument for the predicate, using the heuristics described in Section 3. A binary classifier is then applied to the remaining constituents to do further separation. Finally the multicategory classifier is applied to the candidates that the binary classifier passes along. The results of these two experiments are presented in Table 2.

experiments	all			core
	p (%)	r(%)	f(%)	f(%)
constituents known	n/a	n/a	86.6	86.9
constituents unknown	69.7	73.7	71.6	72.0

Table 2: Results for hand-crafted parses

Compared with the 93.9% reported by Xue and Palmer (2005) for verbal predicates on the same data, the 86.9% the system achieved when the consituents are given is considerably lower, suggesting that SRL for nominalized predicates is a much more challenging task. The difference between the SRL accuracy for verbal and nominalized predicates is even greater when the constituents are not given and the system has to identify the arguments to be classified. Xue and Palmer reported an f-score of 91.4% for verbal predicates under similar experimental conditions, in contrast with the 71.6% our system achieved for nominalized predicates. Careful error analysis shows that one important cause for

435

this degradation in performance is the fact that there is insufficient training data for the system to reliably separate support verbs from other verbs and determine whether the constituents outside the NP headed by the nominalized predicate are related to the predicate or not.

4.3 Using automatic parses

We also conducted an experiment that assumes a more realistic scenario in which the input is raw unsegmented text. We use a fully automatic parser that integrates segmentation, POS tagging and parsing. Our parser is similar to (Luo, 2003) and is trained and tested on the same data partition as the semantic role labeling system. Tested on the held-out test data, the labeled precision and recall are 83.06% and 80.15% respectively for all sentences. The results are comparable with those reported in Luo (Luo, 2003), but they cannot be directly compared with most of the results reported in the literature, where correct segmentation is assumed. In addition, in order to account for the differences in segmentation, each character has to be treated as a leaf of the parse tree. This is in contrast with word-based parsers where words are terminals. Since semantic role tagging is performed on the output of the parser, only constituents in the parse tree are candidates. If there is no constituent in the parse tree that shares the same text span with an argument in the manual annotation, the system cannot possibly get a correct annotation. In other words, the best the system can do is to correctly label all arguments that have a constituent with the same text span in the parse tree.

all			core
p (%)	r(%)	f(%)	f(%)
49.7	53.1	51.3	48.3

Table 3: Results for automatic parses

The results show a similar performance degradation compared with the results reported for verbs on the same data in previous work, which is not unexpected. Xue and Palmer (2005) reported an f-score of 61.3% when a parser is used to preprocess the data.

4.4 Using verb data to improve noun SRL accuracy

Since verbs and their nominalized counterparts are generally considered to share the same argument structure and in fact the Chinese Nombank is annotated based on the same set of lexical guidelines (called *frame files*) as the Chinese PropBank, it seems reasonable to expect that adding the verb data to the training set will improve the SRL accuracy of the nominal predicates, especially when the training set is relatively small. Given that verbs and their nominalized counterpart share the same morphological form in Chinese, adding the verb data to the training set is particularly straightforward. In our experiments, we extracted verb instances from the CPB that have nominalized forms in the portion of the Chinese Treebank on which our SRL experiments are performed and added them to the training set. Our experiments show, however, that simply adding the verb data to the training set and indiscriminately extracting the same features from the verb and noun instances will hurt the overall performance instead of improving it. This result is hardly surprising upon closer examination: the values of certain features are vastly different for verbal and nominal predicates. Most notably, the path from the predicate to the constituent being classified, an important feature for semantic role labeling systems, differ greatly from nominal and verbal predicates. When they are thrown in the same training data mix, they effectively create noise and neutralize the discriminative effect of this feature. Other features, such as the head words and their POS tags, are the same and adding these features does indeed improve the SRL accuracy of nominal predicates, although the improvement is not statistically significant.

4.5 Reranking

In a recent paper on the SRL on verbal predicates for English, (Toutanova et al., 2005) pointed out that one potential flaw in a SRL system where each argument is considered on its own is that it does not take advantage of the fact that the arguments (not the adjuncts) of a predicate are subject to the hard constraint that they do not have the same label[3]. They

---

[3]For certain symmetrical predicates, arguments can have the same label, although these cases are rare.

show that by performing joint learning of all the arguments in the same proposition (for the same predicate), the SRL accuracy is improved. To test the efficacy of joint-learning for nominalized predicates in Chinese, we conducted a similar experiment, using a perceptron reranker described in Shen and Joshi (2004). Arguments and adjuncts of the same predicate instance (proposition) are chained together with their joint probability being the product of the individual arguments and the top K propositions are selected as the reranking candidates. When the arguments are given and the input is hand-crafted gold-standard parses in the treebank, selecting the top 10 propositions yields an oracle score of 97%. This initial promise does not pan out, however. Performing reranking on the top 10 propositions did not lead to significant improvement, using the five feature classes described in (Haghighi et al., 2005). These are features that are hard to implement for individual arguments: *core argument label sequence*, *flattened core argument label sequence*, *core argument labels and phrase type sequence*, *repeated core argument labels with phrase types*, *repeated core argument labels with phrase types and adjacency information*. We speculate that the lack of improvement is due to the fact that the constraint that core (numbered) arguments should not have the same semantic role label for Chinese nominalized predicates is not as rigid as it is for English verbs. However further error analysis is needed to substantiate this speculation.

5 Related Work

Compared with large body of work on the SRL of verbal predicates, there has been relatively little work done in analyzing the predicate-argument structure of nominalized predicates. There are even less work done for the nominalized predicates for Chinese. (Hull and Comez, 1996) implemented a rule-based system for identifying the arguments for nominal predicates and (Lapata, 2002) has a system that interprets the relation between the head of noun compound and its head, but no meaningful comparison can be made between our work and theirs. Perhaps the closest work to that of ours is that of (Pradhan et al., 2004a), where they reported preliminary work for analyzing the predicate-argument structure of Chinese nominalizations, using a small data set of

630 proposition for 22 nominalizations taken from the Chinese Treebank. Since different data sets are used, the results cannot be meaningfully compared.

The results reported here for nominalized predicates are consistent with what Xue and Palmer (2005) reported for the SRL of Chinese verbs with regard to the role of the parser in their semantic role labeling system: there is a substantial performance drop when the automatic parser is used. At present, improvement in Chinese parsing is hindered by insufficient training material. Although the Chinese Treebank has a decent size of 500K words, it is evenly divided into two portions of very different sources, Xinhua newswire from mainland China and Sinorama magazines from Taiwan. Due to their very different styles, training on one portion of the data does not help or even hurt the parsing accuracy of the other portion. The lack of sufficient training material is compounded by inherent properties of the Chinese language that makes Chinese parsing particularly difficult. Chinese segmentation is a much more difficult problem than tokenization of English text and Chinese words do not have morphological clues that can help parsing decisions. We believe further improvement in SRL accuracy will be to a large extent contingent on the parsing accuracy, which requires more training material.

6 Conclusion and future work

We reported first results on the semantic role labeling of nominalized predicates in Chinese, using a sizable annotated corpus, the Chinese Nombank, as training and test data. Compared with that of verbal predicates, SRL of nominalized predicates generally presents a more challenging problem, for all experimental conditions. While the smaller training set compared with that of verbal predicates may provide partial explanation for the degradation in performance, we believe another important reason is that the arguments for nominalized predicates do not occupy prominent syntactic positions such as the subject and object, as arguments of verbal predicates often do. As a result, the syntactic structure represented in the parse tree does not provide as much of a clue for their detection and classification. However, this makes SRL of nominalized predicates a more pressing issue to solve, as they represent a substan-

tial proportion of the predicates in the corpus. Our results also show that the k-best propositions produced by the local classifier have a very high oracle score, which perhaps indicates a promising path that deserves further exploration, based on careful analysis of the errors. We intend to continue to experiment with new features and parameters for the reranking algorithm.

7 Acknowledgement

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of COLING/ACL*, pages 86–90, Montreal, Canada.

Xavier Carreras and Lluís Màrquez. 2004a. Hierarchical Recognition of Propositional Arguments with Perceptrons. In *Proceedings of the Eighth Conference on Natural Language Learning*, Boston, Massachusetts.

Xavier Carreras and Lluís Màrquez. 2004b. Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling. In *Proceedings of the Eighth Conference on Natural Language Learning*, Boston, Massachusetts.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In *Proceedings of the Nineth Conference on Natural Language Learning*, Ann Arbor, Michigan.

D. Gildea and D. Jurafsky. 2002. Automatic labeling for semantic roles. *Computational Linguistics*, 28(3):245–288.

Kadri Hacioglu, Sameer Pradhan, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2003. Shallow Semantic Parsing Using Support Vector Machines. Technical Report CSLR-2003-1, Center for Spoken Language Research at the University of Colorado.

Aria Haghighi, Kristina Toutanova, and Christopher Manning. 2005. A Joint Model for Semantic Role Labeling. In *Proceedings of the Nineth Conference on Natural Language Learning*, Ann Arbor, Michigan.

Richard D. Hull and Fernando Comez. 1996. Semantic interpretation of nominalizations. In *The AAAI Conference*, pages 1062–1068, Oregon.

Peter Koomen, Vasin Punyakanok, Dan Roth, and Wen tau Yih. 2005. Generalized Inference with Multiple Semantic Role Labeling Systems. In *Proceedings of the Nineth Conference on Natural Language Learning*, Ann Arbor, Michigan.

Maria Lapata. 2002. The disambiguation of nominalizations. *Computational Linguistics*, 28(3):357–388.

Xiaoqiang Luo. 2003. A Maximum Entropy Chinese Character-Based Parser. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, Sapporo, Japan.

A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The Nom-Bank Project: An Interim Report. In *Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*, Boston, Massachusetts.

Martha Palmer, Dan Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1).

Sameer Pradhan, Honglin Sun, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2004a. Parsing Arguments of Nominalizations in English and Chinese. In *Proceedings of NAACL-HLT 2004*, Boston, Mass.

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, and Daniel Jurafsky. 2004b. Shallow Semantic Parsing Using Support Vector Machines. In *Proceedings of NAACL-HLT 2004*, Boston, Mass.

Libin Shen and Aravind K. Joshi. 2004. Flexible Margin Selection for Reranking with Full Pairwise Samples. In *Proceedings of IJCNLP-2004*, pages 446–455.

Honglin Sun and Daniel Jurafsky. 2004. Shallow Semantic Parsing of Chinese. In *Proceedings of NAACL 2004*, Boston, USA.

Kristina Toutanova, Aria Haghighi, and Christopher Manning. 2005. Joint Learning Improves Semantic Role Labeling. In *Proceedings of ACL-2005*.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for Semantic Role Labeling. In *Proceedings of 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.

Nianwen Xue and Martha Palmer. 2005. Automatic Semantic Role Labeling for Chinese verbs. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh, Scotland.

Nianwen Xue. 2006. Annotating the predicate-argument structure of Chinese nominalizations. In *Proceedings of the fifth international conference on Language Resources and Evaluation*, Genoa, Italy.