Automatic Semantic Role Labeling for Chinese Verbs

Nianwen Xue, Martha Palmer
CIS Department
University of Pennsylvania
Philadelphia, PA 19104, USA
{xueniwen,mpalmer}@linc.cis.upenn.edu

Abstract

Recent years have seen a revived interst in semantic parsing by applying statistical and machine-learning methods to semantically annotated corpora such as the FrameNet and the Proposition Bank. So far much of the research has been focused on English due to the lack of semantically annotated resources in other languages. In this paper, we report first results on semantic role labeling using a pre-release version of the Chinese Proposition Bank. Since the Chinese Proposition Bank is superimposed on top of the Chinese Treebank, i.e., the semantic role labels are assigned to constituents in a treebank parse tree, we start by reporting results on experiments using the hand-crafted parses in the treebank. This will give us a measure of the extent to which the semantic role labels can be bootstrapped from the syntactic annotation in the treebank. We will then report experiments using a fully automatic Chinese parser that integrates word segmentation, POS-tagging and parsing. This will gauge how successful semantic role labeling can be done for Chinese in realistic situations. We show that our results using hand-crafted parses are slightly higher than the results reported for the state-of-the-art semantic role labeling systems for English using the Penn English Proposition Bank data, even though the Chinese Proposition Bank is smaller in size. When an automatic parser is used, however, the accuracy of our system is much lower than the English state-of-the-art. This reveals an interesting cross-linguistic difference between the two languages, which we attempt to explain. We also describe a method to induce verb classes from the Proposition Bank "frame files" that can be used to improve semantic role labeling.

1 Introduction

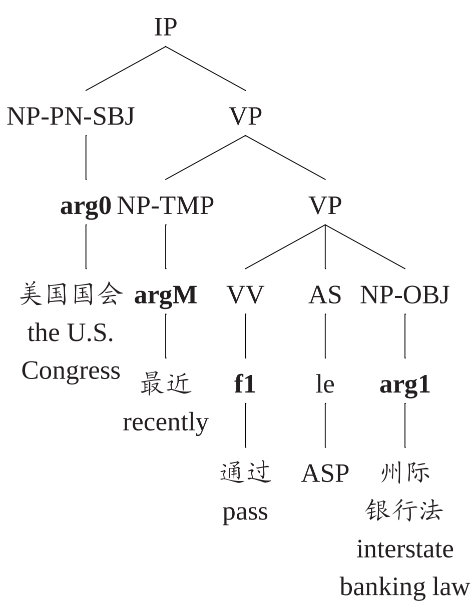
Recent efforts on semantic annotation have made it possible to train domain-independent semantic systems [Gildea and Jurafsky, 2002; Gildea and Palmer, 2002; Hacioglu *et al.*, 2003; Sun and Jurafsky, 2004; Pradhan *et al.*, 2004; Xue and Palmer, 2004]. Most of the semantic annotation projects focus on the predicate-argument structure, which represents a predicate and a number of arguments that are expected of this predicate. Generally each expected argument is assigned a label that marks the role this argument plays in relation to its predicate. It is in the level of generalization these role labels represent that the various annotation efforts differ. The most general are a limited set of roles such as agent and theme that are globally meaningful [Chen *et al.*, 2004]. The role labels used in FrameNet [Baker *et al.*, 1998] are less general in that they are meaningful only with respect to a specific situation, more formally known as a semantic *frame*. For example, the label *Byr* is only meaningful in the "Commercial_transaction" frame. One reflection of this reduced generality is that it is realized with a small class of predicates that indicate transaction, e.g. *purchase*, *rent*. The least general are the labels used in the Propbank annotation. The Propbanks [Palmer *et al.*, 2005; Xue and Palmer, 2003] use predicate-specific labels *ARG0*, *ARG1*, ... *ARGn* for arguments and *ARGM* combined with a secondary tag to mark adjunct-like elements. The secondary tags indicate types of adjuncts and represent generalizations across all verbs.

The predicate-specific approach of the Propbank annotation builds a solid foundation for making high-level generalizations in a bottom-up manner, if broader generalizations are needed. There is generally a straightforward mapping from the numbered role labels to more general roles such as agent and theme. It is much harder to derive these semantic concepts from syntactic representation because an argument may not always be realized, and when it is, it may not always be realized in the same syntactic position as a result of syntactic alternations [Levin, 1993], etc.. In addition, different senses of a verb take different sets of arguments that demonstrate different syntactic patterns. Thus, predicate-argument structure recognition at this level represents a crucial leap towards proper representation of semantic structure from the syntactic structure.

So far, most of the work on automatic semantic role labeling has been devoted to English due to the lack of resources in other languages. In this paper we present first results on the Chinese Proposition Bank [Xue and Palmer, 2003]. While Sun and Jurafsky [2004] did preliminary work on Chinese semantic role labeling using 10 selected verbs, our experiments will be on the all the verbs in the Chinese Proposition Bank. We discuss the linguistic annotation of the Chinese Proposition Bank in Section 2. We describe our experiments using hand-crafted parses in Section 3. In particular, we describe a method of automatically inducing verb classes and use this information as features in semantic role labeling. In Section 4, we describe our experiments using a fully automatic Character-based parser. Section 5 discusses our experimental results and the facilitating and exacerbating factors in Chinese semantic role labeling in comparison with English. Section 6 concludes this paper.

2 The Chinese Proposition Bank

In this section we briefly examine the annotation scheme of the Penn Chinese Propbank [Xue and Palmer, 2003]. The Chinese Propbank is based on the Chinese Treebank [Xue *et al.*, To apear], which is a 500K-word corpus annotated with syntactic structures. The semantic annotation in the Propbank is added to the appropriate constituents in a syntactic tree. This is illustrated in Table 1. The newly added semantic role labels are in bold. The predicate 通过"pass" in Table 1 has two numbered arguments, *ARG0* 美国国会"U.S. Congress" and *ARG1* 州际银行法"interstate banking law". It also has a temporal semantic adjunct, *ARGM-TMP* 最近"recently". In addition to the arguments and adjuncts, the predicate 通过"pass" is also annotated with a *frameset* identifier, *f1*, which uniquely identifies the frameset of this verb instance.



The U.S. Congress recently passed the interstate banking law.

Table 1: A treebank tree annotated with semantic role labels and frameset ID

The frameset represents a major sense defined by the set of arguments a predicate takes. For example, (1) and (2) represents two different framesets of the predicate 通过"pass" because its arguments in those two cases are different. The two arguments for frameset *f1* can be described as "a legislative body" and "the law, regulations, etc. that are passed", while the two arguments for frameset *f2* are "an entity in motion" and "a location the entity passes through". Note that the two instances of 通过"pass" within (1) or (2) do not belong to different framesets since they share the same set of arguments, even though the arguments may be realized in different places, if realized at all. In (1b), the "legislative body" argument is dropped and the "law" argument is realized in the subject position instead of the object position (1a).

(1) a. [**arg0** 美国 国会] 最近 [**f1** 通过] 了 [**arg1** 州际 银行法] 。
the U.S. Congress recently pass ASP interstate banking law .

"The U.S. Congress passed the interstate banking law recently."

b. [**arg1** 州际 银行法] 最近 [**f1** 通过] 了 。
interstate banking law recently pass ASP .

"The interstate banking law passed recently."

(2) a. [**arg0** 火车] 正在[**f2** 通过] [**arg1** 隧道] 。
train now pass tunnel .

"The train is passing through the tunnel."

b. [**arg1** 火车] 正在[**f2** 通过] 。
train 正在 pass .

"The train is passing through."

The task of semantic role labeling is to use the role labels as categories and classify each argument as belonging to one of these categories.

3 Semantic role tagging with hand-crafted parses

In this section we describe a system that does semantic role labeling 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 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.

3.1 Classifier

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.

3.2 Architecture

The Propbank annotation is predicate-centered in that sense that only constituents that are semantic arguments and adjuncts (in a loose sense) to a predicate are annotated. Since

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

the treebank sentences are very long and generally contain several verbs, the majority of the constituents are not related to the predicate in question. One obvious strategy is to assign a *NULL label* to the unannotated constituents, but it is a known fact that when negative samples (*NULL* constituents in this case) overwhelm positive samples (constituents that are actually annotated), the classifier will be heavily biased towards *NULL* constituents. Most systems find a way to filter out some of the negative samples to make the classification task more balanced. For example, [Hacioglu *et al.*, 2003] uses a two-stage architecture where a binary classifier is first used to label all the constituents as either *NULL* or *NON-NULL*, and then a multi-category classifier is run on the *NON-NULL* constituents to assign the semantic role labels. Xue and Palmer [2004] uses a three-stage architecture in which some negative samples are first filtered out with heuristics that exploit the syntactic structures represented in a parse tree. A binary classifier is then applied to further separate the positive samples from the negative samples and finally a multiply-category classified is applied to assign the semantic role labels to the positive samples. In our experiments we adopt the same three-stage strategy described in [Xue and Palmer, 2004]. First the heuristic algorithm is adapted to take advantage of the syntactic structures in the Penn Chinese Treebank, which uses a different set of syntactic categories. The algorithm starts from the predicate that anchors the annotation, and first collects all the syntactic complements of this predicate, which are represented as sisters to the predicate. It then iteratively moves one level up to the parent of the current node till it reaches the appropriate top-level node. This top-level node can be an IP, CP, or NP depending on what their parent nodes are. For example, if the parent node of a CP is an NP, then this CP is very likely a relative clause and the head of this NP is a possible argument to the predicate, via a trace. In this case the algorithm will collect all sisters of this CP. 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. It is worth pointing out that the functional tags and traces are not used to determine the candidates to allow for fair comparison with results from using automatic parses, where they are not available. 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 then negative samples to maximize the recall so that we can pass along as many as positive samples as possible to the final stage, which is the multi-category classification.

3.3 Data

In all our experiments we use a pre-release version of the Chinese Proposition Bank. This version of the Chinese Proposition Bank [Xue and Palmer, 2003] 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 the data has 250K words and 10,364 sentences. The total number of verb types in this chunk of the data is 4,854.[3]. Following the convention of the English semantic role labeling experiments, we divide the training and test data by the number of articles, not by the verb instances. This pretty much guarantees that there will be unseen verbs in the test data. For all our experiments on semantic role labeling, 661 files (`chtb_100.fid` to `chtb_931.fid`) are used as training data and the other 99 files (`chtb_001.fid` to `chtb_099.fid`) are held out as test data. However, our parser is trained on all the data in most current version of the Penn Chinese Treebank except for the test data that has been set aside. That is, in addition to the training data for the semantic role labeling experiments, it also uses the rest of the treebank which has not been propbanked.

3.4 Features

One characteristic of feature-based semantic role modeling is that the feature space is generally large. This is in contrast with the low-level NLP tasks such as POS tagging, which generally have a small feature space. A wide range of features have been shown to be useful in previous work on semantic role labeling [Gildea and Jurafsky, 2002; Pradhan *et al.*, 2004; Xue and Palmer, 2004] and we suspect that many more will be tested before they will settle down with a core set of features. In their preliminary work on Chinese semantic role labeling, Sun and Jurafsky [2004] has successfully adapted a number of the features to Chinese. In our experiments more features that have been described in recent work on English semantic role labeling are adapted to Chinese. We briefly discuss these features and explain at an intuitive level why these are useful features for semantic role labeling where necessary. Our focus, however, will be on how verb classes can be induced from "frame files" from the Penn Proposition Bank and be used as features, which we will discuss in greater detail in the next subsection. The features that we use are listed below:

- *Position*: The position is defined in relation to the predicate verb and the values are *before* and *after*.
- *path*: The path between the constituent in focus and the predicate.
- *Head word and its part of speech*: The head word and its part-of-speech is often a good indicator of the semantic role label of a constituent.
- *Predicate*: The verb itself.
- *subcat frame*: The rule that expands the parent of the verb.
- *Phrase type*: The syntactic category of the constituent in focus.
- *First and last word of the constituent in focus*
- *Phrase type of the sibling to the left*
- *syntactic frame*: The syntactic frame consists of the NPs that surround the predicate verb. This feature is defined

[2]The most current version (CTB5.0) of the Penn Chinese Treebank has 507K words, 825K Chinese characters, 18,716 sentences and 890 articles.

[3]These include the so-called stative verbs, which roughly correspond to adjectives in English.

by the position of the constituent in focus in relation to this syntactic frame [Xue and Palmer, 2004].

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

The reason why the position feature is useful is obvious since constituents receiving a particular semantic role label may occur in some typical positions. For example, the majority of the adjuncts, *ARGM*s occur before the verb in Chinese. The path feature, defined as the route from the constituent in focus to the predicate, represents a more "fine-grained" position. While the values for the simple position feature are just "BEFORE" or "AFTER", the values for the path feature can represent syntactic notions like "subject" or "object". For example, a subject may be represented as "NP↑IP↓VP↓VV" and an object may be represented as "VV↑VP↓NP". Intuitively the path feature is more informative than simple position features but they are also sparse because they are more specific.

The head word and its part-of-speech are clearly informative for semantic role labeling. For example, a noun phrase headed by "jintian/today" is very likely to be a temporal element. So is a prepositional phrase with the head word "zai/at". However, for propositional phrases, the preposition is not always the most informative element. Sometimes the head word of the NP complement is more predictive of the semantic category. For example, in the prepositional phrase "zai/at beijing/Beijing", the NP head "beijing/Beijing" is more telling of the fact that it indicates a location. So for prepositional phrases we use both the preposition and the head noun as features in our system. As has been discussed in Sun and Jurafsky, the head word feature also tends to be sparse, especially given the smaller size of the Penn Chinese Treebank. The chance of seeing a word in the test data that also occurs in the training data is small. The POS tag serves as one form of backoff: constituents headed by words that have the same part-of-speech are likely to receive the same semantic role labels as well.

The first-word-in-the-constituent and the phrase label of the left sibling features are from [Pradhan *et al.*, 2004] and the interested reader is referred to their work for an explanation of why these are useful features. We also implemented features that are described in Xue and Palmer [2004]. These include the syntactic template features, predicate head word combination features and predicate phrase type combination features. For details of these features the reader is referred to [Xue and Palmer, 2004].

3.5 Using verb classes to improve semantic role labeling

With the current experiment setup, as is also the case in most of the work on semantic role labeling, training data and testing data are not divided by verb instances but by the number of articles. As a result, it is expected that the verb instances are not evenly divided. It is even possible that some verbs can only be found in the training data and other verbs can only be found in the test data. By our count, there are 4,666 verb types in the training data and 1,067 verb types in the test data. 188 verb types that occur in the test data are absent from the training data and conversely, 3,787 verb types are in the training data only. Since the semantic role labels are defined with regard to the individual verbs, this can be a real problem since the model learned in the training process does not optimally fit with the test data if different verbs are involved. Fortunately, many verbs have similar argument structures and therefore are annotated with similar semantic role labels in the Chinese Proposition Bank. For examples, verbs like "jiada/enlarge", "jiaju/make more drastic", "jiakuai/accelerate", "jiaqiang/strengthen", "jiashen/deepen", "jiasu/accelerate", "jiazhong/give more weight" "jiagao/make higher" all take two arguments, a theme that undergoes a change of state and an external force or agent that brings about the change of state. These verbs are uniformly annotated and they all have two numbered arguments *arg0* and *arg1* with *arg0* denoting the cause and *arg1* denoting the theme. It would make sense to group these verbs together into a class and use this information as features in the semantic role labeling task. Having a membership in a particular class says something about the predicate-argument structure of a verb and when a verb is absent in the training data, which is the familiar sparse data problem, the class information may tell the system how to label the semantic roles of the verbs belonging to a particular class.

Although to our knowledge no such classification exists for Chinese verbs based on the predicate-argument structures, a rough classification can be automatically derived from the frames files, which are created to guide the Propbank annotation. We classified the verbs along three dimensions, the number of arguments, the number of framesets and selected syntactic alternations.

Number of arguments: Verbs in the Chinese Proposition Bank can have one to five arguments, with the majority of them having one, two or three arguments. Verbs with zero argument are auxiliary verbs [4] like "bi/will", "deyi/be able to", "gai/should", "gan/dare", "ke/may", "ken/be willing to", "neng/can", "nenggou/can", "xu/must", "yingdang/should" and some other light verbs. Verbs that have five arguments are change of state verbs like "yanchang/lengthen", "suoduan/shorten" "jiangdi/lower", "tigao/increase", "kuoda/enlarge", "suoxiao/make smaller". These verbs generally take as arguments a theme that undergoes the change of state, the original state, the new state, the range of the change and the cause or agent that brings about the change.

Number of framesets: A frameset roughly corresponds to a major sense. This information is used because it is common that the different framesets of a verb can have different number of arguments. For example, verbs like "pingheng/balance" can be used either as a non-stative verb, in which case it means "balance", or a stative verb, in which case it means "balanced". When it is used an a non-stative verb, it takes two arguments, the thing or situation that is balanced and the balancer, the entity that

[4] One could say that the argument of the auxiliary verbs is the entire proposition, but in this phase of the Chinese Proposition Bank, auxiliary verbs are not annotated.

maintains the balance. When it is used as a static verb, obviously it only takes a single argument.

Syntactic alternations: We also represent certain type of syntactic alternations. One salient type of syntactic alternation is the well-known "subject of intransitive / object of the transitive" alternation described in detail in Levin [Levin, 1993]. Chinese verbs that demonstrate this alternation pattern include "chuban/publish". For example, "zhe/this ben/CL shu/book" plays the same semantic role even if it is the subject in "zhe/this ben/CL shu/book chuban/publish le/AS" and the object in "zhe/this jia/CL chubanshe/publishing house chuban/publish le/ASP zhe/this ben/CL shu/book".

Thus each verb will belong to a class with a symbol representing each of the three dimensions. For example, a given verb may belong to the class "C1C2a", which means that this verb has two framesets, with the first frameset having one argument and the second having two arguments. The "a" in the second frameset represents a type of syntactic alternation. 40 classes are derived in this manner.

Such a classification scheme will undoubtedly prove to be linguistically unsophisticated. Verbs that have the same number of arguments may have different types of arguments, and the current classifications system do not pick up these distinctions. However, our experiments show that such a simple classification, when used as features, significantly improves the semantic role labeling task.

4 Using automatic parses

Previous work [Sun and Jurafsky, 2004] on Chinese semantic role labeling uses a parser that assume correct (hand-crafted) segmentation. As word segmentation is a very challenging problem that has attracted a large body of research by itself, it is still unclear how well semantic role tagging in Chinese can be performed in realistic situations. In our experiments, we use a fully automatic parser that integrates segmentation, POS tagging and parsing. Our parser is similar to [Luo, 2003]. The parser is trained on CTB5.0, using the training data described in the previous section. Tested on the held-out test data, the labeled precision and recall are 81.83% and 82.91% 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.

5 Results and Discussion

The results of the semantic role labeling are presented in Table 2. The verb class information, when used as features, improves the semantic role labeling accuracy by about 1 percent when the Gold Standard parses are used. When the parser is used, however, the improvement is insignificant. The 93.9 percent that our system achieved using hand-crafted parses is fairly high considering the fact that the state-of-the-art semantic role labeling systems trained on the English Proposition Bank [Palmer *et al.*, 2005] is around 93 percent [Pradhan *et al.*, 2004; Xue and Palmer, 2004] and the English Proposition Bank is a much larger corpus, with 1 million words. There are several facilitating factors for Chinese semantic role labeling when hand-craft parses are provided as input. First of all, Chinese verbs appear to be less polysemous, at least the ones that occur in the Penn Chinese Treebank. Of the 4854 verbs in this version of the Chinese Proposition bank, only 62 verbs have 3 or more framesets. In contrast, 294 verbs out of the 3635 verbs in the Penn English Proposition Bank has 3 or more framesets. When a verb is less polysemous, the arguments of the verb tend to be realized in a more uniform manner in syntax. As a result, the argument labels are easier to predict from their structure in a correct parse tree. Chinese seems to compensate this fact by using a larger number of verbs. This becomes obvious when we consider the fact that 4854 verbs are from just 250K words and the 3635 verbs in the English Proposition Bank is from one million words! A related fact is that adjectives in Chinese are traditionally counted as verbs and they generally have only one argument with a much simpler syntactic realization.

experiments	class	p (%)	r(%)	f(%)
known constituents (Gold)	no	n/a	n/a	92.7
known constituents (Gold)	yes	n/a	n/a	93.9
unknown constituents (Gold)	no	90.4	90.3	90.3
unknown constituents (Gold)	yes	91.4	91.1	91.3
unknown constituents (parser)	no	66.1	57.0	61.2
unknown constituents (parser)	yes	67.0	56.4	61.3

Table 2: Results

We also believe that a more subtle explanation for the higher semantic role labeling accuracy given the annotation of the Chinese Treebank is the fact that the Chinese Treebank has richer structure. By using less flat structures and more hierarchical structures, the Chinese Treebank resolves some of the attachment ambiguities that are useful for semantic role labeling. For example, the complement and adjunct in a VP in the Chinese Treebank are attached in different syntactic configurations with regard to the verb. Since complements are generally numbered arguments and adjuncts are generally *ARGM*s, the semantic role labeler can take advantage of this fact it tries to determine when a constituent is a numbered argument or an adjunct.

This apparent advantage in Chinese semantic role labeling is diminished when an automatic parser is used. First of all, the hierarchical structures in the hand-craft parses that aid semantic role labeling are hard to recover with an automatic parser. Resolving the many attachment ambiguities caused by the hierarchical structures in language is the one of the most difficult problems in parsing literature. Parsing Chinese in a realistic scenario is especially difficult given that word segmentation must be performed as a prerequisite. Chinese word segmentation is a non-trivial problem that has been the topic of a lot of research in the Chinese language processing community. It is still a very much unsolved problem, with the current state-of-the-art accuracy being around 96 percent [Sproat and Emerson, 2003]. This is in contrast with the much less difficult problem of English tokenization. The segmentation errors are very difficult to recover in later parsing stages. Another property of Chinese that exacerbates parsing is the lack of formal clues. For example, determining whether a word is a verb or noun in Chinese can be difficult because Chinese does not have a rich inflection system to aid POS tagging.

6 Conclusions and Future Work

We presented first experimental results on Chinese semantic role labeling using a pre-release version of the Chinese Proposition Bank. We show that verb classes, induced from the predicate-argument information in the "frame files", helps semantic role labeling. We also show that given Gold Standard parses, Chinese semantic role labeling can be performed with considerable accuracy. In fact, even though the Chinese Proposition Bank is a smaller corpus the English Proposition Bank, we achieved results that are slightly higher than the state-of-the-art English semantic role labeling systems. On the other hand, even though our experiments using a fully automatic parser yields promising results, the accuracy is much lower than the English state-of-the-art. We suggest that the lower accuracy when an automatic parser is used is due to the fact that determining the argument boundaries in Chinese using a parser is a very challenging task because word segmentation is a prerequisite and word segmentation errors are hard to recover in the parsing stage. The lack of formal clues such as inflection morphology also makes POS tagging a very difficult task. Since parsing assumes accurate word segmentation and POS tagging, word segmentation and POS tagging errors naturally translate into parsing errors. Once these bumps are overcome, however, semantic analysis actually gets easier, compared with English, due to the lower degree of polysemy.

Naturally there is a lot to be done in Chinese semantic role labeling, with the first order of business being to improve Chinese parsing. We also need to conduct more experiments with the features to figure out which features are most useful for Chinese.

Acknowledgements

We would like to thank Zixin Jiang, Meiyu Chang and Gang Chen for their annotation efforts. This work is funded by the NSF ITR via grant 130-1303-4-541984-XXXX-2000-1070.

References

[Baker *et al.*, 1998] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of COLING/ACL*, pages 86–90, Montreal, Canada, 1998.

[Chen *et al.*, 2004] Keh-Jiann Chen, Chu-Ren Huang, Feng-Yi Chen, Chi-Ching Luo, Ming-Chung Chang, and Chao-Jan Chen. Sinica Treebank: Design Criteria, Representational Issues and Implementation. In Anne Abeillé, editor, *Building and Using Parsed Corpora*. Kluwer, 2004.

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

[Gildea and Palmer, 2002] Dan Gildea and Martha Palmer. The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*, Philadelphia, PA, 2002.

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

[Levin, 1993] Beth Levin. *English Verbs and Alternations: A Preliminary Investigation*. Chicago: The Unversity of Chicago Press, 1993.

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

[Palmer *et al.*, 2005] Martha Palmer, Dan Gildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 2005.

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

[Sproat and Emerson, 2003] Richard Sproat and Tom Emerson. The First International Chinese Word Segmentation Bakeoff. In *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, Sapporo, Japan, 2003.

[Sun and Jurafsky, 2004] Honglin Sun and Daniel Jurafsky. Shallow semantic parsing of chinese. In *Proceedings of NAACL 2004*, Boston, USA, 2004.

[Xue and Palmer, 2003] Nianwen Xue and Martha Palmer. Annotating the Propositions in the Penn Chinese Treebank. In *The Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing*, Sapporo, Japan, 2003.

[Xue and Palmer, 2004] Nianwen Xue and Martha Palmer. Calibrating features for semantic role labeling. In *Proceedings of 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, 2004.

[Xue *et al.*, To apear] Nianwen Xue, Fei Xia, Fu dong Chiou, and Martha Palmer. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, To apear.