# Chinese Semantic Role Labeling with Shallow Parsing

**Weiwei Sun** and **Zhifang Sui** and **Meng Wang** and **Xin Wang**

Institute of Computational Linguistics

Peking University

Key Laboratory of Computational Linguistics

Ministry of Education, China

weiwsun@gmail.com;{szf,wm}@pku.edu.cn;xinwang.cpku@gmail.com;

## Abstract

Most existing systems for Chinese Semantic Role Labeling (SRL) make use of full syntactic parses. In this paper, we evaluate SRL methods that take partial parses as inputs. We first extend the study on Chinese shallow parsing presented in (Chen et al., 2006) by raising a set of additional features. On the basis of our shallow parser, we implement SRL systems which cast SRL as the classification of syntactic chunks with IOB2 representation for semantic roles (i.e. semantic chunks). Two labeling strategies are presented: 1) directly tagging semantic chunks in one-stage, and 2) identifying argument boundaries as a chunking task and labeling their semantic types as a classification task. For both methods, we present encouraging results, achieving significant improvements over the best reported SRL performance in the literature. Additionally, we put forward a rule-based algorithm to automatically acquire Chinese verb formation, which is empirically shown to enhance SRL.

## 1 Introduction

In the last few years, there has been an increasing interest in Semantic Role Labeling (SRL) on several languages, which consists of recognizing arguments involved by predicates of a given sentence and labeling their semantic types. Nearly all previous Chinese SRL research took full syntactic parsing as a necessary pre-processing step, such as (Sun and Jurafsky, 2004; Xue, 2008; Ding and Chang, 2008). Many features are extracted to encode the complex syntactic information. In English SRL research, there have been some attempts at relaxing the necessity of using full syntactic parses; better understanding of SRL with shallow parsing is achieved by CoNLL-2004 shared task (Carreras and Màrquez, 2004). However, it is still unknown how these methods perform on other languages, such as Chinese.

To date, the best SRL performance reported on the Chinese Proposition Bank (CPB) corresponds to a F-measure is 92.0, when using the hand-crafted parse trees from Chinese Penn Treebank (CTB). This performance drops to 71.9 when a real parser is used instead[1] (Xue, 2008). Comparatively, the best English SRL results reported drops from 91.2 (Pradhan et al., 2008) to 80.56 (Surdeanu et al., 2007). These results suggest that as still in its infancy stage, Chinese full parsing acts as a central bottleneck that severely limits our ability to solve Chinese SRL. On the contrary, Chinese shallow parsing has gained a promising result (Chen et al., 2006); hence it is an alternative choice for Chinese SRL.

This paper addresses the Chinese SRL problem on the basis of shallow syntactic information at the level of phrase chunks. We first extend the study on Chinese chunking presented in (Chen et al., 2006) by raising a set of additional features. The new set of features yield improvement over the strong chunking system described in (Chen et al., 2006). On the basis of our shallow parser, we implement lightweight systems which solve SRL as a sequence labeling problem. This is accomplished by casting SRL as the classification of syntactic chunks (e.g. NP-chunk) into one of semantic labels with IOB2 representation (**?**). With respect to the labeling strategy, we distinguish two different approaches. The first one directly recognizes semantic roles by an IOB-type sequence tagging. The second approach divides the problem into two independent subtasks: 1) Argument Identification (AI) and 2) Semantic Role Classification (SRC).

---

[1]This F-measure is evaluated on the basis of hand-crafted word segmentation and POS tagging.

A Chinese word consists of one or more characters, and each character, in most cases, is a morpheme. The problem of how the words are constructed from morphemes, known as word formation, is very important for a majority of Chinese language processing tasks. To capture Chinese verb formation information, we introduce a rule-based algorithm with a number of heuristics. Experimental results indicate that word formation features can help both shallow parsing and SRL.

We present encouraging SRL results on CPB[2]. The best F-measure performance (74.12) with gold segmentation and POS tagging can be achieved by the first method. This result yield significant improvement over the best reported SRL performance (71.9) in the literature (Xue, 2008). The best recall performance (71.50) can be achieved by the second method. This result is also much higher than the best reported recall (65.6) in (Xue, 2008).

## 2 Related Work

Previous work on Chinese SRL mainly focused on how to implement SRL methods which are successful on English, such as (Sun and Jurafsky, 2004; Xue and Palmer, 2005; Xue, 2008; Ding and Chang, 2008). Sun and Jurafsky (2004) did the preliminary work on Chinese SRL without any large semantically annotated corpus of Chinese. Their experiments were evaluated only on ten specified verbs with a small collection of Chinese sentences. This work made the first attempt on Chinese SRL and produced promising results. After the CPB was built, (Xue and Palmer, 2005) and (Xue, 2008) have produced more complete and systematic research on Chinese SRL. Ding and Chang (2008) divided SRC into two sub-tasks in sequence. Under the hierarchical architecture, each argument should first be determined whether it is a core argument or an adjunct, and then be classified into fine-grained categories. Chen et al. (2008) introduced an application of transductive SVM in Chinese SRL. Because their experiments took hand-crafted syntactic trees as input, how transductive SVMs perform in Chinese SRL in realistic situations is still unknown.

Most existing systems for automatic Chinese SRL make use of a full syntactic parse of the sentence in order to define argument boundaries and

to extract relevant information for training classifiers to disambiguate between role labels. On the contrary, in English SRL research, there have been some attempts at relaxing the necessity of using syntactic information derived from full parse trees. For example, Hacioglu and Ward (2003) considered SRL as a chunking task; Pradhan et al. (2005) introduced a new procedure to incorporate SRL results predicted respectively on full and shallow syntactic parses. Previous work on English suggests that even good labeling performance has been achieved by full parse based SRL systems, partial parse based SRL systems can still enhance their performance. Though better understanding of SRL with shallow parsing on English is achieved by CoNLL-2004 shared task (Carreras and Màrquez, 2004), little is known about how these SRL methods perform on Chinese.

## 3 Chinese Shallow Parsing

There have been some research on Chinese shallow parsing, and a variety of chunk definitions have been proposed. However, most of these studies did not provide sufficient detail. In our system, we use chunk definition presented in (Chen et al., 2006), which provided a chunk extraction tool. The tool to extract chunks from CTB was developed by modifying the English tool used in CoNLL-2000 shared task, Chunklink[3], and is publicly available at http://www.nlplab.cn/chenwl/chunking.html. The definition of syntactic chunks is illustrated in Line *CH* in Figure 1. For example, "保险公司/the insurance company", consisting of two nouns, is a noun phrase.

With IOB2 representation (Ramshaw and Marcus, 1995), the problem of Chinese chunking can be regarded as a sequence labeling task. In this paper, we first implement the chunking method described in (Chen et al., 2006) as a strong baseline. To conveniently illustrate, we denote a word in focus with a fixed window $w_{-2}w_{-1}ww_{+1}w_{+2}$, where $w$ is current token. The baseline features includes:

- Uni-gram word/POS tag feature: $w_{-2}$, $w_{-1}$, $w$, $w_{+1}$, $w_{+2}$;

- Bi-gram word/POS tag feature: $w_{-2\_}w_{-1}$, $w_{-1\_}w$, $w_{\_}w_{+1}$, $w_{+1\_}w_{+2}$;

---

| WORD: | 截止 | 目前 | 保险 | 公司 | 已 | 为 | 三峡 | 工程 | 提供 | 保险 | 服务 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| POS: | [P] | [NT] | [NN | NN] | [AD] | [P] | [NR] | [NN] | [VP] | [NN | NN] |
| CH: | [PP | NP] | [NP] | | [ADVP] | [PP | NP | NP ] | [VP] | [NP] | |
| M1: | B-A* | I-A*[4] | B-A0 | | B-AM-ADV | B-A2 | I-A2 | I-A2 | B-V | B-A1 | |
| M2-AI: | B-A | I-A | B-A | | B-A | B-A | I-A | I-A | B-V | B-A | |
| M2-SRC: | AM-TMP | | A0 | | AM-ADV | | A2 | | Rel | A1 | |

Until now, the insurance company has provided insurance services for the Sanxia Project.

Figure 1: An example from Chinese PropBank.

That means 18 features are used to represent a given token. For instance, the bi-gram Word features at 5th word position ("公司/company") in Figure 1 are "，_保险", "保险_公司", "公司_已", "已_为".

To improve shallow parsing, we raised an additional set of features. We will discuss these features in section 5.

# 4 SRL with Shallow Parsing

The CPB is a project to add predicate-argument relations to the syntactic trees of the CTB. Similar to English PropBank, the semantic arguments of a predicate are labeled with a contiguous sequence of integers, in the form of A$N$ (i.e. Arg$N$); the adjuncts are annotated as such with the label AM (i.e. ArgM) followed by a secondary tag that represents the semantic classification of the adjunct. The assignment of argument labels is illustrated in Figure 1, where the predicate is the verb "提供/provide". For example, the noun phrase "保险公司/the insurance company" is labeled as *A0*, meaning that it is the *proto-Agent* of 提供; the preposition phrase "截止目前/until now" is labeled as *AM-TMP*, indicating a temporal component.

## 4.1 System Architecture

SRL is a complex task which has to be decomposed into a number of simpler decisions and tagging schemes in order to be addressed by learning techniques. Regarding the labeling strategy, we can distinguish at least two different strategies. The first one consists of performing role identification directly as IOB-type sequence tagging. The second approach consists of dividing the problem into two independent subtasks.

### 4.1.1 One-stage Strategy

In the one-stage strategy, on the basis of syntactic chunks, we define semantic chunks which do not overlap nor embed using IOB2 representation. Syntactic chunks outside a chunk receive the tag O. For syntactic chunks forming a chunk of type A*, the first chunk receives the *B-A*< tag (Begin), and the remaining ones receive the tag *I-A** (Inside). Then a SRL system can work directly by using sequence tagging techinique. Since the semantic annotation in the PropBank corpus does not have any embedded structure, there is no loss of information in this representation. The line *M1* in Figure 1 illustrates this semantic chunk definition.

### 4.1.2 Two-stage Strategy

In the two-stage architecture, we divide Chinese SRL into two subtasks: 1) semantic chunking for AI, in which the argument boundaries are predicted, and 2) classification for SRC, in which the already recognized arguments are assigned role labels. In the first stage, we define semantic chunks *B-A* which means begin of an argument and *I-A* which means inside of an argument. In the second stage, we solve SRC problem as a multi-class classification. The lines *M2-AI* and *M2-SRC* in Figure 1 illustrate this two-stage architecture. For example, the noun phrase "保险公司/the insurance company" is *proto-Agent*, and thus should be labeled as *B-A* in the AI chunking phase, and then be tagged as *A0*. The phrase "为三峡工程/for the Sanxia Project" consists of three chunks, which should be labeled as *B-A*, *I-A*, and *I-A* respectively in the AI chunking phase, then these three chunks as a whole argument should be recognized as *A2*.

### 4.1.3 Chunk-by-Chunk

There is also another semantic chunk definition, where the basic components of a semantic chunk are words rather than syntactic chunks. A good election for this problem is chunk-by-chunk pro-

---

[4]The semantic chunk labels here are *B-AM-TMP* and *I-AM-TMP*. Limited to the document length, we cannot put all detailed chunk labels in one line in Figure 1.

cessing instead of word-by-word. The motivation is twofold: 1) phrase boundaries are almost always consistent with argument boundaries; 2) chunk-by-chunk processing is computationally less expensive and allows systems to explore a relatively larger context. This paper performs a chunk-by-chunk processing, but admitting a processing by words within the target verb chunks.

## 4.2 Features

Most of the feature templates are "standard", which have been used in previous SRL research. We give a brief description of "standard" features, but explain our new features in detail.[5]

### 4.2.1 Features for Semantic Chunking

In the semantic chunking tasks, i.e. the one-stage method and the first step in the two-stage method, we use the same set of features. The features are extracted from three types of elements: syntactic chunks, target verbs, links between chunks and target verbs. They are formed making use of words, POS tags and chunks of the sentence. Xue (2008) put forward a rough verb classification where verb classes are automatically derived from the frame files, which are verb lexicon for the CPB annotation. This kind of verb class information has been shown very useful for Chinese SRL. Our system also includes this feature. In our experiments, we represent a verb in two dimensions: 1) number of arguments, and 2) number of framesets. For example, a verb may belong to the class "C1C2," which means that this verb has two framesets, with the first frameset having one argument and the second having two arguments.

To conveniently illustrate, we denote a token chunk with a fixed context $w_{i-1}[_{c_k}w_i...w_h...w_j]w_{j+1}$, where $w_h$ is the head word of this chunk $c_k$. The complete list of features is listed here.

#### Extraction on Syntactic Chunks

*Chunk type*: $c_k$.

*Length*: the number of words in a chunk.

*Head word/POS tag*. The rules described in (Sun and Jurafsky, 2004) are used to extract head word.

*IOB chunk tag of head word*: chunk tag of head word with IOB2 representation (e.g. B-NP, I-NP).

*Chunk words/POS tags context*. Chunk context includes one word before and one word after: $w_{i-1}$ and $w_{j+1}$.

*POS tag chain*: sequential containers of each word's POS tag: $w_i\_..._w_j$. For example, this feature for "保险服务" is "NN_NN".

*Position*: the position of the phrase with respect to the predicate. It has three values as *before*, *after* and *here*.

**Extraction on Target Verbs**  Given a target verb $w_v$ and its context, we extract the following features.

*Predicate, its POS tag, and its verb class*.

*Predicate IOB chunk tag context*: the chain of IOB2 chunk tags centered at the predicate within a window of size -2/+2.

*Predicate POS tag context*: the POS tags of the words that immediately precede and follow the predicate.

*Number of predicates*: the number of predicates in the sentence.

**Extraction on Links**  To capture syntactic properties of links between the chunks and the verbs, we use the following features.

*Path*: a flat path is defined as a chain of base phrases between the token and the predicate. At both ends, the chain is terminated with the POS tags of the predicate and the headword of the token.

*Distance*: we have two notions of distance. The first is the distance of the token from the predicate as a number of base phrases, and the second is the same distance as the number of VP chunks.

**Combining Features**  We also combine above features as some new features.

*Conjunctions of position and head word, target verb, and verb class*, including: $position\_w_h$, $position\_w_v$, $position\_w_h\_w_v$, $position\_class$, and $position\_w_h\_class$.

*Conjunctions of position and POS tag of head word, target verb, and verb class*, including: $position\_w_h\_w_v$, $position\_w_h$, and $position\_w_h\_class$.

### 4.2.2 Features for SRC

In the SRC stage of the two-stage method, different from previous work, our system only uses word-based features, i.e. features extracted from words and POS tags, to represent a given argument. Experiments show that a good semantic

---

[5]The source code of our system also provides lots of comments for implementation of all features.

role classifier can be trained by using only word-based features. To gather all argument position information predicted in AI stage, we design a coarse frame feature, which is a sequential collection of arguments. So far, we do not know the detailed semantic type of each argument, and we use *XP* as each item in the frame. To distinguish the argument in focus, we use a special symbol to indicate the corresponding frame item. For instance, the Frame feature for argument 保险服务 is *XP+XP+XP+XP+V+!XP*, where *!XP* means that it is the argument in focus.

Denote 1) a given argument $w_{i-2}w_{i-1}[w_iw_{i+1}...w_{j-1}w_j]w_{j+1}w_{j+2}$, and 2) a given predicate $w_v$. The features for SRC are listed as follows.

*Words/POS tags context of arguments*: the contents and POS tags of the following words: $w_i$, $w_{i-1}$, $w_{i-2}$, $w_{i+1}$, $w_{i+2}$, $w_j$, $w_{j+1}$, $w_{j-1}$, $w_{j-2}$, $w_{j+1}$; the POS tags of the following words: $w_{i+1}$, $w_{i+2}$, $w_{j+1}$, $w_{j+2}$.

*Token Position.*

*Predicate, its POS, and its verb class.*

*Coarse Frame.*

*Combining features*: conjunctions of boundary words, including $w_{i-1\_}w_{j+1}$ and $w_{i-2\_}w_{j+2}$; conjunction of POS tags of boundary words, including $w_{i-1\_}w_{j+1}$ and $w_{i-2\_}w_{j+2}$; conjunction of token position, boundary words, and predicate word, including $position\_w_i\_w_j$, $w_i\_w_j\_w_v$; $position\_w_i\_w_j\_w_v$; conjunction of token position, boundary words' POS tags, and predicate word, also including $position\_w_i\_w_j$, $w_i\_w_j\_w_v$; $position\_w_i\_w_j\_w_v$; conjunction of predicate and frame; conjunction of target verb class and frame; conjunction of boundary words' POS tags, and predicate word.

## 5 Automatic Chinese Verb Formation Analyzing

### 5.1 Introduction to Chinese Word Formation

Chinese words consist of one or more characters, and each character, in most cases, is a morpheme which is the smallest meaningful unit of the language. According to the number of morphemes, the words can be grouped into two sets, simple words (consisting of one morpheme) and compound words (consisting of two morphemes or more). There are 9 kinds of word formation in Chinese compound words, and table 1 shows the detail with examples. Note that, *attributive-head*

and *complementarity* are not for Chinese verbs.

| Types | Examples |
|---|---|
| reduplication | 看看(look) 想想(think) |
| affixation | 激化(intensify) 觉着(feel) |
| subject-verb | 耳闻(hear) 口述(dictate) |
| verb-object | 戒烟(quit smoking) |
| | 理发(haircut) |
| verb-complement | 通知(inform) 栽培(plant) |
| verb-result | 超出(exceed) 煮沸(boil) |
| adverbial-head | 隐居(retreat) 误用(misuse) |
| coordinate | 爱惜(cherish) 追逐(chase) |
| attributive-head* | 谣言(rumor) 医院(hospital) |
| complementarity* | 纸张(paper) 马匹(horse) |

Table 1: Example Words with Formation

The internal structure of a word constraints its external grammatical behavior, and the formation of a verb can provide very important information for Chinese SRL. Take ”超出/exceed” as an example, the two characters are both verbal morphemes, and the character ”超” means ”pass” and the character ”出” with the meaning of ”over” shows the complement of the action of ”超”. In this word, ”出” is usually collocated with an object, and hence a *Patient* role should comes after the verb ”超出”. Note that, the verb ”超”, however, is unlikely to have an object. Take ”理发/haircut” as another example, the first character ”理” is a verbal morpheme with the meaning of ”cut” and the second character ”发” is a nominal morpheme with the meaning of ”hair”. In this word, ”发” acts as the object of ”理”, and the word ”理发” is unlikely to have an *Patient* any more in the sentential context.

### 5.2 Verb Formation Analyzing Method

To automatically analyze verb formation, we introduce a rule-based algorithm. Pseudo code in Algorithm 1 illustrates our algorithm. This algorithm takes three string (one or more Chinese characters) sets as *lexicon* knowledge:

- adverbial suffix set $\mathcal{A}$: strings in $\mathcal{A}$ are usually realized as the modifier in a *adverbial-head* type word, e.g. 不/not, 别/not, 总/always, 并/both, 共/all.

- object head set $\mathcal{O}$: strings in $\mathcal{O}$ are usually realized as the head in a *verb-object* type word, e.g. 变/change, 获/get, 谈/talk, 发/send.

---

**Algorithm 1**: Verb Formation Analyzing.

**Data**: adverbial suffix set $\mathcal{A}$, object head set $\mathcal{O}$, complement suffix set $\mathcal{C}$

**input** : word $W = c_1...c_n$ and its POS $P$

**output**: head character $h$, adverbial character $a$, complement character $c$, object character $o$

**begin**

    $h = c = a = o = null$;

    **if** $n = 4$ **and** $c_1 = c_3$ **and** $c_2 = c_4$ **then**

        | **return** *Verb formation of* $W' = c_1c_3$;

    **else if** $n = 3$ **and** $c_2 = c_3$ **then**

        | $h = c_1, c = c_2$;

    **else if** $n = 2$ **and** $c_1 = c_2$ **then**

        | $h = c_1$;

    **else if** $n = 1$ **then**

        | $h = c_1$;

    **else if** $c_n \in \mathcal{C}$ **and** $c_{n-1}c_n \in \mathcal{C}$ **and** $P="VV"$ **then**

        | $h = c_1, c = c_n/c_{n-1}c_n$;

    **else if** $c_1 \in \mathcal{A}$ **then**

        | $a = c_1, h = c_2...c_n$;

    **else if** $c_1 \in \mathcal{O}$ **and** $P="VV"$ **then**

        | $h = c_1, o = c_2...c_n$;

**end**

---

- complement suffix set $\mathcal{C}$: strings in $\mathcal{C}$ are usually realized as complement in a verb-complement type word: e.g. 出/out, 入/in, 完/finish, 来/come, 不到/not.

Note that, to date there is no word formation annotation corpus, so direct evaluation of our rule-based algorithm is impossible. This paper makes task-oriented evaluation which measures improvements in SRL.

### 5.3 Using Word Formation Information to improve Shallow Parsing

The majority of Chinese nouns are of type *attributive-head*. This means that for most nouns the last character provides very important information indicating the head of the noun. For example, the word formations of "桃树/peach", "柳树/willow" and "黄杨树/boxtree" (three different kinds of trees), are *attributive-head* and they have the same head word "树/tree". While for verbs, the majority are of three types: *verb-object*, *coordinate* and *adverbial-head*. For example, words "加大/enlarge", "加剧/make more drastic" and "加快/accelerate" have the same *head* "加/add". The *head* morpheme is very useful in alleviating the

data sparseness in word level. However, for any given word, it is very hard to accurately find the head. In the shallow paring experiments, we use a very simple rule to get a *pseudo* head character: 1) extracting the last word for a noun, and 2) extracting the first word for a verb. The new features include:

*Pattern 1*: conjunction of pseudo head of $w_{i-1}$ and POS tags of $w_{i-1}$ and $w_i$.

*Pattern 2*: conjunction of pseudo head of $w_i$ and POS tags of $w_{i-1}$ and $w_i$.

*Pattern 3*: conjunction of length/POS tags of $w_{i-1}, w_i, w_{i+1}$.

### 5.4 Using Verb Formation Information to improve SRL

We use some new verb formation features to improve our SRL system. The new features are listed as follows. The first four are used in semantic chunking task, and all are used in SRC task.

*First/last characters*.

*Word length*.

*Conjunction of word length and first/last character*.

*Conjunction of token position and first/last character*.

*The head string* of a verb (e.g. "理" in "理发").

*The adverbial string* of a verb (e.g. "误" in "误用").

*The complement string* of a verb (e.g. "出" in "超出").

*The object string* of a verb (e.g. "发" in "理发").

## 6 Results and Discussion

### 6.1 Experimental Setting

#### 6.1.1 Data

Experiments in previous work are mainly based on CPB and CTB, but the experimental data preparing procedure does not seem consistent. For example, the sum of each semantic role reported in (Ding and Chang, 2008) is extremely smaller than the corresponding occurrence statistics in original data files in CPB. In this paper, we modify CoNLL-2005 shared task software[6] to process CPB and CTB. In our experiments, we use the CPB 1.0 and CTB 5.0. The data is divided into three parts: files from chtb_081 to chtb_899 are used as training set; files from chtb_041 to

---

[6]http://www.lsi.upc.edu/~srlconll/soft.html

chtb_080 as development set; files from chtb_001 to chtb_040, and chtb_900 to chtb_931 as test set. The data setting is the same as (Xue, 2008). The results were evaluated for precision, recall and F-measure numbers using the srl-eval.pl script provided by CoNLL-2005 shared task.

### 6.1.2 Classifier

For both syntactic and semantic chunking, we used TinySVM along with YamCha[7] (Kudo and Matsumoto, 2000; Kudo and Matsumoto, 2001). In the chunking experiments, all SVM classifiers were realized with a polynomial kernel of degree 2. Pair-wise strategy is used to solve multi-class classification problem. For the SRC experiments, we use a linear SVM classifier, along with One-Vs-All approach for multi-class classification. $SVM_{lin}$[8], a fast linear SVM solvers, is used for supervised learning. $l_2$-SVM-MFN (modified finite newton) method is used to solve the optimization problem (Keerthi and DeCoste, 2005).

### 6.2 Shallow Parsing Performance

|  | P(%) | R(%) | $F_{\beta=1}$ |
|---|---|---|---|
| Baseline | 93.54 | 93.00 | 93.27 |
| Ours | 93.83 | 93.39 | 93.61 |

Table 2: Shallow parsing performance

Table 2 summarizes the overall shallow parsing performance on test set. The first line shows the performance of baseline. Comparing the best system performance 94.13 F-measure of CoNLL 2000 shared task (Syntactic Chunking on English), we can see Chinese shallow parsing has reached a comparable result, tough the comparison of numeric performance is not very fair, because of different languages, different chunk definition, different training data sizes, etc.. The second line *Ours* shows the performance when new features are added, from which we can see the word formation based features can help shallow parsing.

Table 3 shows the detailed performance of noun phrase (NP) and verb phrase (VP), which make up most of phrase chunks in Chinese. Our new features help NP more, whereas the effect of new features for VP is not significant. That is in part because most VP chunk recognition error is caused by long dependency, where word formation fea-

---
[7]http://chasen.org/~taku/index.html.en

[8]http://people.cs.uchicago.edu/~vikass/svmlin.html

|  | P(%) | R(%) | $F_{\beta=1}$ |
|---|---|---|---|
| NP(Baseline) | 90.84 | 90.05 | 90.44 |
| NP(Ours) | 91.42 | 90.78 | 91.10 |
| VP(Baseline) | 94.44 | 94.55 | 94.50 |
| VP(Ours) | 94.65 | 94.74 | 94.69 |

Table 3: Performance of NP-chunk and VP-chunk

tures do not work. Take the sentences below for example:

1. [$_{VP}$ 因此获得胜利]。 (Therefore (we) achieve victory.)

2. [$_{ADVP}$ 因此] [$_{VP}$ 大量出现] 的是以前不曾遇到的。 (Therefore the major changes have not been met before.)

The contexts of the word "因此/therefore" in the two sentences are similar, where "因此" is followed by verbal components. In the second sentence, the word "因此/therefore" will be correctly recognized as an adverbial phrase unless classifier knows the following component is a clause. Unfortunately, word formation features cannot supply this kind of information.

### 6.3 SRL Performance

|  | P(%) | R(%) | A(%) | $F_{\beta=1}$ |
|---|---|---|---|---|
| (Xue, 2008) | 79.5 | 65.6 | – | 71.9 |
| M1− | 79.02 | 69.12 | – | 73.74 |
| M1+ | 79.25 | 69.61 | – | 74.12 |
| M2−/AI | 80.34 | 75.11 | – | 77.63 |
| M2+/AI | 80.01 | 75.15 | – | 77.51 |
| M2−/SRC | – | – | 92.57 | – |
| M2+wf/SRC | – | – | 93.25 | – |
| M2+/SRC | – | – | 93.42 | – |
| M2−AI+SRC | 76.48 | 71.50 | – | 73.90 |

Table 4: Overall SRL performance of different methods

Table 4 lists the overall SRL performance numbers on test set using different methods mentioned earlier; these results are based on features computed from gold standard segmentation and POS tagging, but automatic recognized chunks, which is parsed by our improved shallow parsing system. For the AI and the whole SRL tasks, we report the precision (P), recall (R) and the $F_{\beta=1}$-measure scores, and for the SRC task we report the classification accuracy (A). The first line *(Xue,*

1481

*2008)* shows the SRL performance reported in (Xue, 2008). To the authors' knowledge, this result is best SRL performance in the literature. Line 2 and 3 shows the performance of the one-stage systems: 1) Line *M1−* is the performance without word formation features; 2) Line *M1+* is the performance when verb formation features are added. Line 4 to 8 shows the performance of the two-stage systems: 1) Line *M2−/AI* and *M2+/AI* shows the performance of AI phase without and within word formation features respectively; 2) Line *M2−/SRC* shows the SRC performance with trivial word-based features (i.e. frame features and verb formation features are not used); 3) Line *M2+wf/SRC* is the improved SRC performance when coarse verb formation features are added; 4) Line *M2+/SRC* is the SRC performance with all features; 5) Line *M2−AI+SRC* shows the performance of SRL system, which uses baseline features to identify arguments, and use all features to classify arguments.

## 6.4 Discussion

The results summarized in Table 4 indicate that according to the-state-of-the-art in Chinese parsing, SRL systems based on shallow parsing outperforms the ones based on full parsing. Comparison between *one-stage strategy* and *two-stage strategy* indicates 1) that there is no significant difference in the F-measure; and 2) that *two-stage strategy* method can achieve higher recall while *one-stage strategy* method can achieve higher precision. Both the *one-stage strategy* and *two-stage strategy* methods yield significant improvements over the best reported SRL performance in the literature, especially in terms of recall performance. Comparison SRL performance with full parses and partial parses indicates that both models have strong and weak points. The full parse based method can implement high precision SRL systems, while the partial parse based methods can implement high recall SRL systems. This is further justification for combination strategies that combine these independent SRL models.

Generally, Table 4 shows that verb formation features can enhance Chinese SRL, especially for fine-grained role classification. The effect of word formation in formation in both shallow parsing and SRL suggests that automatic word formation analyzing is very important for Chinese language processing. The rule-based algorithm is just a preliminary study on this new topic, which requires

| Num of words | P (%) | R (%) | $F_{\beta=1}$ |
|---|---|---|---|
| Length = 1 | 84.69% | 75.48% | 79.82 |
| Length = 2 | 82.14% | 74.21% | 77.97 |
| Length = 3 | 75.43% | 63.98% | 69.24 |
| Length = 4 | 75.71% | 65.63% | 70.32 |
| Length = 5 | 72.46% | 64.38% | 68.18 |
| Length = 6 | 72.97% | 66.21% | 69.43 |
| Length = 7 | 77.03% | 67.65% | 72.04 |
| Length = 8 | 74.39% | 57.28% | 64.72 |
| Length = 9 | 66.67% | 51.16% | 57.89 |
| Length = 10 | 68.08% | 58.28% | 62.80 |
| Length = 11+ | 67.40% | 57.71% | 62.18 |

Table 5: SRL performance with arguments of different length

more research effort.

Though our SRC module does not use any parsing information, our system can achieve 93.42% accuracy, comparing the best gold parse based result 94.68% in the literature. This result suggests that Chinese SRC system, even without parsing, can reach a considerable good performance. The main reason is that in Chinese, arguments with different semantic types have discriminative boundary words, which can be extracted without parsing. It is very clear that the main bottleneck for Chinese SRL is to accurately identify arguments rather than to disambiguate their detailed semantic types.

Table 5 summarizes the labeling performance for argument of different length. It is not surprising that arguments are more and more difficult to rightly recognize as the increase of their length. But the performance decline slows up when the length of arguments is larger than 10. In other words, some of the arguments that are composed of many words can still be rightly identified. The main reason for this point is that these arguments usually have clear collocation words locating at argument boundaries. Take the sentences below for example,

3. 包括[*A1* ......等] (including ... etc.)

the object of the verb "包括/include" has a definite collocation word "等/etc.", and therefore this object is easy to be recognized as a *A1*.

## 7 Conclusion

In this paper, we discuss Chinese SRL on the basis of partial syntactic structure. Our systems advance the state-of-the-art in Chinese SRL. We first

extend the study on Chinese shallow parsing and implement a good shallow parser. On the basis of partial parses, SRL are formulated as a sequence labeling problem, performing IOB2 decisions on the syntactic chunks of the sentence. We exploit a wide variety of features based on words, POS tags, and partial syntax. Additionally, we discuss a language special problem, i.e. Chinese word formation. Experimental results show that coarse word formation information can help shallow parsing, especially for NP-chunk recognition. A rule-based algorithm is put forward to automatically acquire Chinese verb formation, which is empirically shown to enhance SRL.

## Acknowledgments

## References

Xavier Carreras and Lluís Màrquez. 2004. Introduction to the conll-2004 shared task: Semantic role labeling. In Hwee Tou Ng and Ellen Riloff, editors, *HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, pages 89–97, Boston, Massachusetts, USA, May 6 - May 7. Association for Computational Linguistics.

Wenliang Chen, Yujie Zhang, and Hitoshi Isahara. 2006. An empirical study of Chinese chunking. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 97–104, Sydney, Australia, July. Association for Computational Linguistics.

Yaodong Chen, Ting Wang, Huowang Chen, and Xishan Xu. 2008. Semantic role labeling of Chinese using transductive svm and semantic heuristics. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*.

Weiwei Ding and Baobao Chang. 2008. Improving Chinese semantic role classification with hierarchical feature selection strategy. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 324–333, Honolulu, Hawaii, October. Association for Computational Linguistics.

Kadri Hacioglu and Wayne Ward. 2003. Target word detection and semantic role chunking using support vector machines. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 25–27, Morristown, NJ, USA. Association for Computational Linguistics.

S. Sathiya Keerthi and Dennis DeCoste. 2005. A modified finite newton method for fast solution of large scale linear svms. *J. Mach. Learn. Res.*, 6:341–361.

Taku Kudo and Yuji Matsumoto. 2000. Use of support vector learning for chunk identification. In *Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning*, pages 142–144, Morristown, NJ, USA. Association for Computational Linguistics.

Taku Kudo and Yuji Matsumoto. 2001. Chunking with support vector machines. In *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001*, pages 1–8, Morristown, NJ, USA. Association for Computational Linguistics.

Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Semantic role chunking combining complementary syntactic views. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 217–220, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Sameer S. Pradhan, Wayne Ward, and James H. Martin. 2008. Towards robust semantic role labeling. *Comput. Linguist.*, 34(2):289–310.

L. A. Ramshaw and M. P. Marcus. 1995. Text chunking using transformation-based learning. In *Proceedings of the 3rd ACL/SIGDAT Workshop on Very Large Corpora, Cambridge, Massachusetts, USA*, pages 82–94.

Honglin Sun and Daniel Jurafsky. 2004. Shallow semantc parsing of Chinese. In Daniel Marcu Susan Dumais and Salim Roukos, editors, *HLT-NAACL 2004: Main Proceedings*.

Mihai Surdeanu, Lluís Màrquez, Xavier Carreras, and Pere Comas. 2007. Combination strategies for semantic role labeling. *J. Artif. Intell. Res. (JAIR)*, 29:105–151.

Nianwen Xue and Martha Palmer. 2005. Automatic semantic role labeling for Chinese verbs. In *in Proceedings of the 19th International Joint Conference on Artificial Intelligence*, page 2005.

Nianwen Xue. 2008. Labeling chinese predicates with semantic roles. *Computational Linguistics*, 34(2):225–255.