Deep Context-Free Grammar for Chinese with Broad-Coverage

Xiangli Wang
Japan Patent Information Organization, Tokyo, Japan
xiangli_wang@japio.or.jp

Yi Zhang
Dept of Computational Linguistics and DFKI GmbH, Saarland University, Saarland, Germany
yizhang@dfki.de

Yusuke Miyao
National Institute of Informatics, Tokyo, Japan
yusuke@nii.ac.jp

Takuya Matsuzaki
National Institute of Informatics, Tokyo, Japan
takuya-matsuzaki@nii.ac.jp

Junichi Tsujii
Microsoft Research Asia, Beijing, China
jtsujii@microsoft.com

Abstract

The accuracy of Chinese parsers trained on Penn Chinese Treebank is evidently lower than that of the English parsers trained on Penn Treebank. It is plausible that the essential reason is the lack of surface syntactic constraints in Chinese. In this paper, we present evidences to show that strict deep syntactic constraints exist in Chinese sentences and such constraints cannot be effectively described with context-free phrase structure rules as in the Penn Chinese Treebank annotation; we show that such constraints may be described precisely by the idea of Sentence Structure Grammar; we introduce how to develop a broad-coverage rule-based grammar for Chinese based on this idea; we evaluated the grammar and the evaluation results show that the coverage of the current grammar is 94.2%.

1 Introduction

Penn Treebank (PTB) was built based on the idea of context-free PSG (Marcus et al., 1993). It is now a common practice to develop data-driven English parsers using PTB annotation and encouraging performances have been reported (Collins, 2000; Charniak, 2000).

Following the success of PTB, Xue et al. 2000 built Penn Chinese Treebank (CTB). CTB is also based on context-free PSG. Since CTB provides training data for Chinese parsing, researchers attempted to train Chinese parsing with CTB (Bikel and Chiang, 2000; Chiang and Bikel, 2002; Levy and Manning, 2003; Bikel, 2004; Wang et al., 2006; Zhang and Clark, 2009; Huang et al., 2009). However, these works showed that the performances of Chinese parsing were significantly worse than English.

Such inferior performances can be the result of several factors. One of them being that Chinese is an isolating language. Verbs and nouns of Chinese have little morphological paradigms so that the surface syntactic constraints of Chinese sentences less than English sentences. For example, the word "process" acts as different roles in English sentences 1a), 1b) and 1c). The morphologies of the word provide constraints for the roles that it acts as. As a contrast, "处理/process" acts as different roles also in Chinese sentences 2a), 2b) and 2c), but there is no morphology change of the word. Either English PSG rules of PTB or Chinese PSG rules of CTB describe surface syntactic structures of sentences. The lack of surface syntactic constraints of Chinese causes that PSG rules of CTB for Chinese sentences are looser than PSG rules of PTB for English sentences. Therefore, we speculate that the lack of surface syntactic constraints of Chinese sentences is the essential reason why the performances of Chinese PSG parsing are lower than English obviously.

1a. Students process data
1b. Data processing system
1c. Data was processed

2a. 学生 处理 数据
Student process data
Students process data
2b. 数据 处理 系统
Data process system
Data processing system

Proceedings of the Seventh SIGHAN Workshop on Chinese Language Processing (SIGHAN-7), pages 11–19,
Nagoya, Japan, 14 October 2013.

2c. 数据　处理　了
Data　process　le
Data was processed

There is another question: are there strict deep syntactic constraints in Chinese sentences? If there were strict deep syntactic constraints in Chinese sentences, and there was grammar formulism capable of describing such constraints precisely, then it would be possible to further improve the performances of Chinese parsing.

In this paper, we present evidences to show that there are strict deep syntactic constraints in Chinese sentences, which are constraints of co-occurrence between deep sentence structures and predicate verbs, but such constraints cannot be described with PSG rules of CTB (section 2); we present examples to show that the idea of Sentence Structure Grammar (SSG) can describe such deep syntactic constraints so that SSG rules can analyze Chinese sentences deeper and more precisely than PSG rules of CTB (section 3);we also show how a broad-coverage Chinese grammar was developed based on SSG (section 4); we evaluate the coverage of the grammar and the results show that its coverage is satisfactory (section 5).

2 Deep Syntactic Constraints in Chinese Sentences

There are plenty of evidences showing that strict deep syntactic constraints exist in Chinese sentences. These are constraints of co-occurrence between deep sentence structures and predicate verbs. We present some examples here.

Sentences (3a-3c) and (4a-4c) can be abstracted into two deep structures: 5a) and 5b). Since the structures like 5a) and 5b) describe the relations between the predicate and its semantic-related constituents like "Agent" and "Direction", we call such structures as ***deep sentence structures***. The deep sentence structures 5a) and 5b) accept "飞/fly" as their predicates but not "吃/eat", and "喜欢/like". Therefore, 3a) and 4a) are grammatical sentences but 3b), 3c), 4b) and 4c) are ungrammatical.

3a. 鸟儿　向　南方　飞
Bird towards south　fly
Birds fly towards the south
3b. *鸟儿　向　南方　吃
Bird towards south　eat
Birds eat towards the south
3c. *鸟儿　向　南方　喜欢
Bird towards south　like
Birds like towards the south
4a. 鸟儿　飞　向　南方
Bird fly towards south
Birds fly towards the south
4b. *鸟儿 吃　向　南方
Bird eat towards south
Birds eat towards the south
4c. *鸟儿 喜欢　向　南方
Bird like towards south
Birds like towards the south
5a. Agent Direction V
5b. Agent V xiang4 Direction

Sentences (6a-6c) and (7a-7c) can be abstracted into two deep sentence structures: 8a) and 8b). 8a) and 8b) accept "吃/eat" as their predicates but not "飞/fly" and "喜欢/like". That is why 6a) and 7a) are grammatical sentences but 6b), 6c), 7b) and 7c) are ungrammatical.

6a. 鸟儿 把 种子 吃 了
Bird ba seed eat le
Birds ate the seeds
6b. *鸟儿 把 种子 飞 了
Bird ba seed fly le
Birds fly the seeds
6c. *鸟儿 把 种子 喜欢 了
Bird ba seed like le
Birds liked the seeds
7a. 种子 被 鸟儿 吃 了
Seed bei bird eat le
Seeds were eaten by birds
7b. *种子 被 鸟儿 飞 了
Seed bei bird fly le
Seeds were flied by birds
7c. *种子 被 鸟儿 喜欢 了
Seed bei bird like le
Seeds were liked by birds

8a. Agent ba Object V le
8b. Object bei Agent V le

Sentences (9a-9c) and (10a-10c) can be abstracted into two deep sentence structures: 11a) and 11b). 11a) and 11b) accept "喜欢/like" as their predicates but not "吃/eat" and "飞/fly". For this reason, the sentences 9a) and 10a) are grammatical but 9b), 9c), 10b) and 10c) are ungrammatical sentences.

9a. 鸟儿 比 狗儿 喜欢 种子
bird than dog like seed
Birds like seeds than dogs
9b. *鸟儿 比 狗儿 飞 种子
bird than dog fly seed
Birds fly seeds than dogs
9c. *鸟儿 比 狗儿 吃 种子
bird than dog eat seed
Birds eat seeds than dogs
10a. 鸟儿 喜欢 狗儿 偷 种子
Bird like dog steal seed
Birds like that dogs steal seeds

10b. *鸟儿 飞 狗儿 偷 种子
Bird fly dogs steal seed
Birds fly that dogs steal seeds
10c. *鸟儿 吃 狗儿 偷 种子
Bird eat dog steal seed
Birds eat that dogs steal seeds

11a. Agent Comparison V Object
11b. Agent V Objects

The above examples provide evidences that deep sentence structures and predicate verbs choose each other. In another words, constraints of co-occurrence between deep sentence structures and predicate verbs exist widely in Chinese sentences.

Deep sentence structures choose predicates according to their deep syntactic properties. "飞/fly" accepts a direction constituent but not an object or a comparison constituent, so it can appear 5a) and 5b) but not 8a), 8b), 11a) and 11b). "吃/eat" accepts an object but not a direction constituent or a comparison constituent, thus it chooses 8a) and 8b) but not 5a), 5b), 11a) and 11b); "喜欢/like" accepts an object, an sentential object or a comparison constituent but not a direction constituent so that it can be predicates of 11a) and 11b) but not 5a), 5b), 8a) and 8b).

Constraints of co-occurrence between deep sentence structures and predicate verbs exist in Chinese sentences commonly. Obviously, CTB rules that describe sentences with context-free phrase structures cannot describe such deep syntactic constraints in Chinese sentences so that distinguish the grammatical sentences from ungrammatical sentences in the above sentences. The rule set of CTB are written to cover the grammatical sentences 3a), 4a), 6a), 7a), 9a), and 10a), but they also cover all ungrammatical sentences above.

12a. IP→ NP-SBJ VP
IP-OBJ→ NP-SBJ VP
VP→ BA IP-OBJ
VP→ LB IP-OBJ
VP→ PP VP
VP→ VP PP
VP→ VV
VP→ VV NP-OBJ
VP→ VV IP-OBJ
PP→ P NP

3 Describing Deep Syntactic Constraints with SSG Rules

Sentence Structure Grammar (SSG) is an idea for grammar formulism (Wang and Miyazaki, 2007; Wang et al., 2012a). SSG focus on describing constraints of co-occurrence between deep sentence structures and predicate verbs that are discussed in section 2. Deep sentence structures in section 2 are treated as rules based on SSG ideas (figure 2); predicate verbs are classified according to their deep syntactic properties (as shown in figure 3); for each type of predicate verbs, only the deep sentence structures that co-occur with them are treated as SSG rules (figure 4). SSG rules not only present deeper information but avoid effectively covering ungrammatical sentences that are covered by CTB rules.

We show how SSG rules present deeper information than CTB rules. SSG is a kind of context-free grammar, but its idea to analyze language is different from context-free PSG. Rather than PSG rules describing a sentence with phrases, SSG rules treat a sentence as a whole that consists of a predicate and its semantic-related constituents. For example, PSG rules of CTB analyze 4a) as shown in figure 1 but SSG rules analyze the same sentence as shown in figure 2. SSG rules present semantic role information like "Agent" and "Direction" besides phrase information such noun phrase, while CTB rules present phrase information and syntactic role like "SBJ".

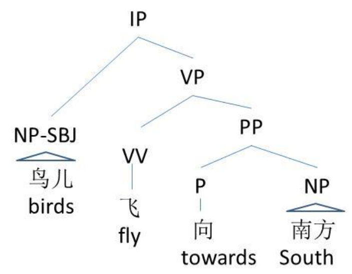


Figure 1: the CTB tree of 4a)

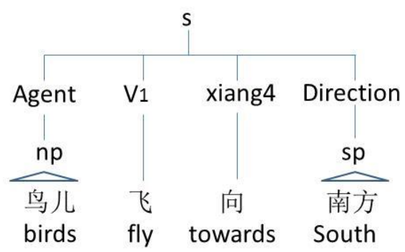


Figure 2: the SSG tree of 4a)

We show how SSG rules avoid covering ungrammatical sentences in section 2, which are covered by CTB rules. Predicate verbs would be classified according to their deep syntactic properties based on SSG ideas. The verbs "飞/fly" belongs to a type that accept an agent and a direction constituent; "吃/eat" belongs to the type that accept an agent and an object but not a direction constituent and a comparison constituent; “喜欢/like” is in a type that accept an agent, an object, a comparison constituent, and a sentential constituent (figure 3).

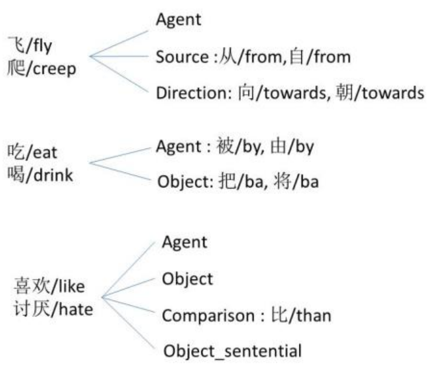


Figure 3: how to classify the predicate verbs based on SSG

For each type of predicate verbs, only deep sentence structures that co-occur with them are treated as rules. As shown in figure 4, for the verbs like “飞/fly”, only 5a) and 5b) are the deep sentence structures that co-occur with them, but 8a), 8b), 11a) and 11b) are not, so only 5a) and 5b) are described as the SSG rules 13a) and 13b) for this type of predicate verbs. In the same way, the deep sentence structures 8a) and 8b) are treated as the SSG rules 14a) and 14b) for the type of predicate verbs like “吃/eat”; the deep sentence structures 11a) and 11b) are written as the SSG rules 15a) and 15b) for the type of predicate verbs like “喜欢/like”. In this way, the SSG rules 13a) and 13b) only cover the grammatical sentences 3a) and 4a) but not cover the ungrammatical sentences 3b), 3c), 4b) and 4c); the SSG rules 14a) and 14b) cover the grammatical sentences 6a) and 7a) but not cover ungrammatical sentences 6b, 6c), 7b) and 7c); the SSG rules 15a) and 15b) cover the grammatical sentences 9a) and 10a) but not cover the ungrammatical sentences 9b), 9c), 10b) and 10c). The constraints of co-occurrence between deep sentence structures and predicate verbs are described precisely by SSG rules by this way.

13a. s→ Agent V1 xiang4 Direction
Agent→ np
Direction→ sp

13b. s → Agent Direction V1
Agent→ np
Direction→ xiang4 sp

14a. s→ Agent ba Object V2 le
Agent→ np
Object→ np

14b. s→ Object bei Agent V2 le
Agent→ np
Object→ np

15a. s→ Agent Comparison V3 Object
Agent→ np
Object→ np
Comparison→ bi3 np

15b. s→ Agent V3 Objects
Agent→ np
Objects→ s
s→ Agent V2 Object

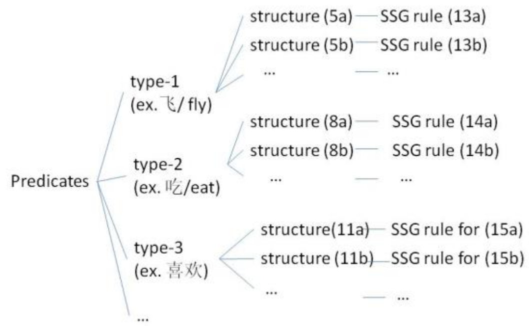


Figure 4: how to develop the SSG rules

4 Grammar Development for Chinese Based on SSG

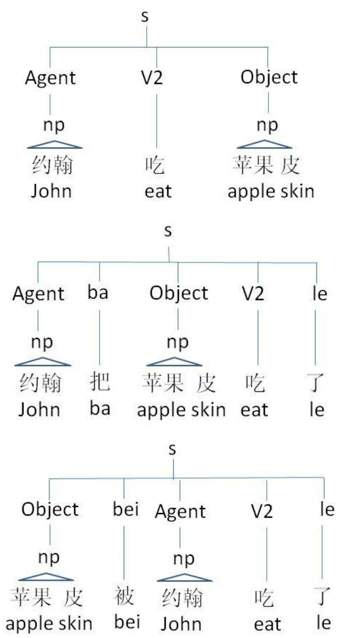
A broad-coverage grammar for Chinese, named Chinese Sentence Structure Grammar (CSSG), had been developed based on SSG (Wang et al., 2012b).

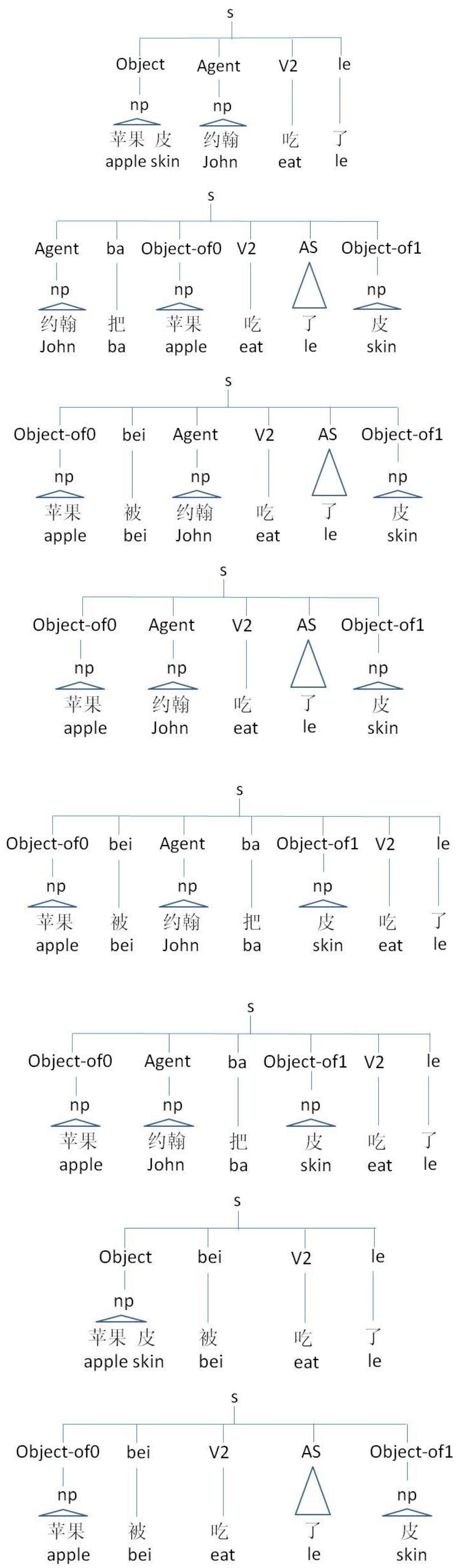
The idea of SSG is helpful for developing broad-coverage grammar. The predicate verbs of Chinese are classified into 52 types according to their deep syntactic properties. Such classification of predicate verbs provides a clear goal for the developer to develop a broad-coverage grammar. It is to cover all deep sentence structures that co-occur with each type of predicate verbs (shown in fig. 4). For example, for the type of predicate verbs like “吃/eat”, the deep sentence structures (16a-16l) are covered by the SSG rules (17a-17l) in CSSG. (16a-16l) include various constructions wide-discussed in linguistic literatures like ba-construction, bei-construction, topic-construction and so on. Figure 5 shows the SSG trees of (16a-16l).

16a. 约翰 吃 苹果 皮
John eat apple skin
John eats apple skin
16b. 约翰 把 苹果 皮 吃 了
John ba apple skin eat le
John ate the apple skin
16c. 苹果 皮 被 约翰 吃 了
apple skin bei John eat le

The apple skin was eaten by John
16d. 苹果 皮 约翰 吃 了
apple skin John eat le
The apple skin was eaten by John
16e. 约翰 把 苹果 吃 了 皮
John ba apple eat le skin
John eats the skin of the apple
16f. 苹果 被 约翰 吃 了 皮
Apple bei John eat le skin
The skin of the apple was eaten by John
16g. 苹果 约翰 吃 了 皮
apple John eat le skin
The skin of the apple was eaten by John
16h. 苹果 被 约翰 把 皮 吃 了
Apple bei John ba skin eat le
The apple skin was eaten by John
16i. 苹果 约翰 把 皮 吃 了
Apple John ba skin eat le
The apples, John ate their skin
16j. 苹果 皮 被 吃 了
Apple skin bei eat le
The apple skin was eaten
16k. 苹果 被 吃 了 皮
Apple bei eat le skin
The apples, their skin was eaten
16l. 苹果 被 把 皮 吃 了
Apple bei ba skin eat le
The apples, their skin was eaten

17a. s→ Agent V2 Object
17b. s→ Agent ba Object V2 le
17c. s→ Object bei Agent V2 le
17d. s→ Object Agent V2 le
17e. s→ Agent ba Object-of0 V2 le Object-of1
17f. s→ Object-of0 bei Agent V2 le Object-of1
17g. s→ Object-of0 Agent V2 le Object-of1
17h. s→ Object-of0 bei Agent ba Object-of1 V2 le
17i. s→ Object-of0 Agent ba Object-of1 V2 le
17j. s→ Object bei V2 le
17k. s→ Object-of0 bei V2 le Object-of1
17l. s→ Object-of0 bei ba Object-of1 V2 le





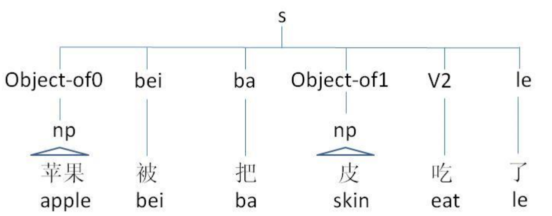


Figure 5: the SSG trees for (16a-16l)

There is a practical issue when developing a broad-coverage grammar based on the SSG idea. It is that the number of SSG rules covering a kind of language would be huge. Wang and Miyazaki 2007 proposed a method to avoid developing a huge number of rules. They divide constituents of a sentence into indispensable parts and dispensable parts. Indispensable constituents must appear while dispensable constituents may or may not appear in a sentence. For example, in the SSG rule set 18a), the asterisked constituents “advp”, “AS” and “y” are dispensable constituents, while “Agent”, “Object” and “V2” are indispensable constituents. By this way, one SSG rule set 18a) can cover a lot of structures, like (19a-19i) (shown in figure 6).

18a. s→ advp* Agent advp* V_2 AS* Object y*
Agent→ np
Object→ np
AS→ le
AS→ zhe
AS→ guo
advp→ tp
advp→ pp-loc

19a. 约翰 吃 了 苹果 皮
John eat le apple skin
John ate the apple skin

19b. 约翰 吃 过 苹果 皮
John eat guo apple skin
John has ever eaten apple skin

19c. 约翰 也 吃 苹果 皮
John also eat apple skin
John eats apple skin also

19d. 约翰 吃 苹果 皮 吗
John eat apple skin ma
Does John eat apple skin

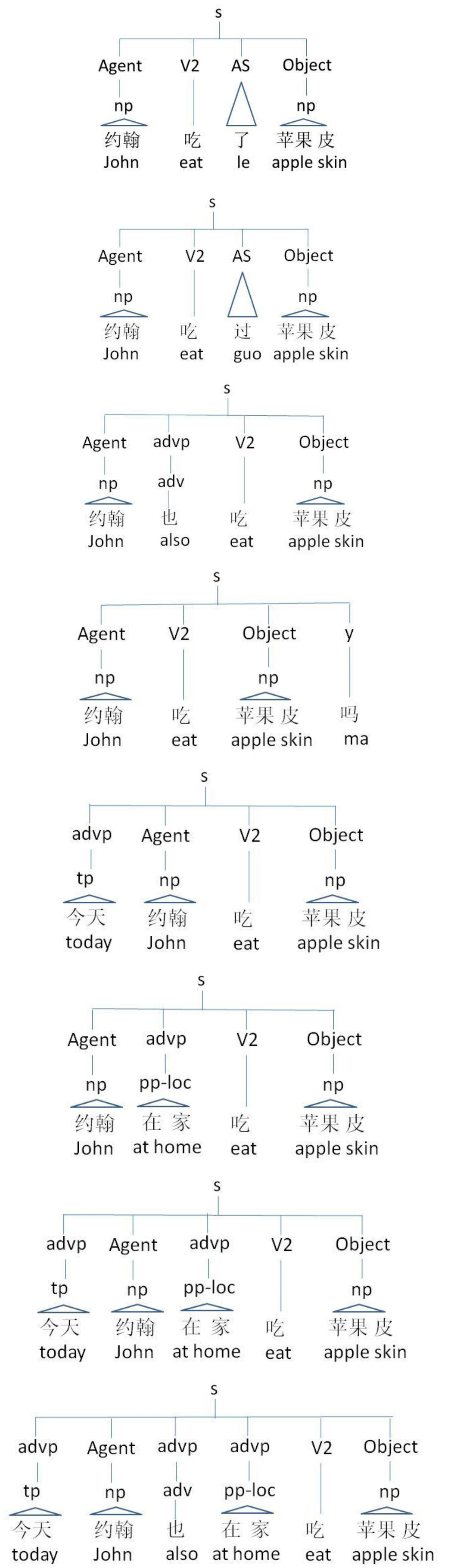
19e. 今天 约翰 吃 苹果 皮
Today John eat apple skin
John eat apple skin today

19f. 约翰 在 家 吃 苹果 皮
Johan at home eat apple skin
John eats apple skin at home

19g. 今天 约翰 在 家 吃 苹果 皮
Today John at home eat apple skin
John eats apple skin at home today

19h. 今天 约翰 也 在 家 吃 苹果 皮
Today John also at home eat apple skin
John also eats apple skin at home today

19i. 今天 约翰 也 在 家 吃 苹果 皮 吗
Today John also at home eat apple skin ma
Does John also eat apple skin at home today



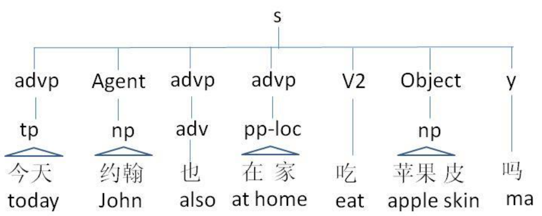


Figure 6: the SSG trees for (19a-19i)

5 Evaluation and Discussion

5.1 Evaluation Results

We evaluated the coverage of CSSG. We chose the first 200 sentences from CTB development data as the test set. We convert the CTB trees of the test data into the CSSG trees semi-automatically with heuristics and some manual correction. Then we evaluate how many constructions of the test data are covered by the CSSG rules.

5,333 construction instances are exacted from the test data (table 1). These may be divided into 3 types:

1) Sentential constructions: the constructions of simple sentences and complex sentences;
2) Semantic roles: the constructions of semantic roles like "Agent", "Object" and "Direction";
3) Phrase constructions: the constructions of phrase like "np", "advp" and "tp".

Among these constructions, 19.1% are the sentential constructions; 14.4% are the semantic roles; 62.9% are the phrase constructions.

Sen. Constr.	Sem. Role	Phr. Constr.	Total
1,014(19.1%)	770(14.4%)	3549(66.5%)	5,333(100%)

Table 1: the contents of the constructions of the test data

	Total	Matched	Unmatched
Sen. Constr.	1014(100%)	905(89.3%)	109(10.7%)
Sem. Role	770(100%)	764(99.2%)	6(0.8%)
Phr. Constr.	3549(100%)	3355(94.5%)	194(5.5%)
Total	5333(100%)	5024(94.2%)	309(5.8%)

Table 2: coverage of the CSSG

Table 2 shows that the coverage of CSSG. 94.2% of the total constructions of the test data are covered by CSSG: 89.3% of sentences constructions; 99.2% of semantic roles; 94.5% of phrase constructions.

Unmatched Sen. Constr.	Unmatched for simple Sen.	Unmatched for complex Sen.
109(100%)	13(11.9%)	96(88.1%)

Table 3: contents of unmatched sentential constructions

Since the coverage of the sentential constructions of the CSSG is lower than the other types, we analyze the unmatched sentential constructions further. As shown in table 3, 88.1% of unmatched sentential constructions are for complex sentences, only 11.9% for simple sentences. 90.5% of the sentential constructions are for simple sentences (table 4) and 98.6% of the constructions for simple sentences are covered by the CSSG (table 5).

Sen. Constr.	Constr. for simple Sen.	Constr. for complex Sen.
1014(100%)	918(90.5%)	96(9.5%)

Table 4: contents of sentential constructions of the test data

Constr. for simple Sen.	Matched	Unmatched
918(100%)	905(98.6%)	13(1.4%)

Table 5: coverage of the simple sentential constructions of CSSG

We analyzed the type of the unmatched constructions for simple sentences. These may be divided into 3 types:

1) The constructions for special structures;
2) The constructions for common structures;
3) The constructions for new types of predicate verbs.

Table 6 summarizes the contents of the unmatched constructions for simple sentences.

the type of unmatched constr.	Number
Special structure	2
Common structure	9
New type of verbs	2
	13

Table 6: analysis of the unmatched constructions for simple sentence

5.2 Discussion

The evaluation results show that the coverage of the sentential constructions of the CSSG is lower than the coverage of the total rules (table 2), but 88.1% of the unmatched constructions are for complex sentences (table 3). As the discussion in section 2 and section 3, the CSSG rules focus on covering the deep sentence structures of simple sentences. The rules for complex sentences are still not included by the current version of the CSSG.

Table 4 shows that 90.5% of the sentential constructions are for simple sentences, and the coverage of the constructions of simple sentences of CSSG is 98.6% (table 5). The results verified that the CSSG rules cover the deep sentence structures of Chinese widely.

There are 13 deep sentence structures that failed to be covered by CSSG (table 5). As shown in table 6, most of them appear commonly but CSSG failed to cover these constructions; two of them are special structures like 20a) and 20b), these structures need to be described with special rules; two of them are not covered because their predicate verbs are not covered by the current version of the CSSG. The two verbs are “获悉/know from” and “符合/be in accord with”. “获悉/know from” accept a sentential object and a source constituent; “符合/be in accord with” accept a nominal subject, a sentential subject and an object (figure 7). These two types of verbs are still not included by the predicate classification of CSSG. It is possible to improve the coverage of CSSG by adding such new types of verbs to the predicate classification of CSSG and describing the SSG rules for them. For example, the predicate verb of 21a) is “获悉/know from”, and 22a) is the deep sentence structure of 21a); the predicate verbs of 23a) and 23b) are “符合/be in accord with”. 24a) and 24b) are the deep sentence structures of 23a) and 23b). We can add the new types of predicates like “获悉/know from” and “符合/be in accord with” to the predicate classification of CSSG, then describe SSG rules for the deep sentence structures 22a), 24a) and 24b). In this way, the coverage of CSSG can be further improved.

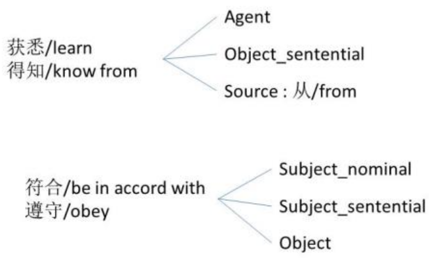


Figure 7: the new types of predicate verbs

20a. 中国 的 友好城市 以 日本 为 最 多
China de sister city yi3 Japan wei2 most few
Japan has most of the sister cities of China

20b. 他 给 人 以 挑战者 的 印象
He give people yi3 challenger de impression
He gives people an impression of a challenger

21a. 他 从 记者 获悉 日本 发生 地震
He from reporter learn Japan happen earthquake
He learned from reporters that there was an earthquake in Japan

22a. Agent Source Vi Object_sentential

23a. 他 符合 雇用条件
He be in accord with employment condition
He is in accord with the employment condition

23b. 减少 工资 符合 公司利益
Decrease salary be in accord with company’s benefit
It is in accord with company’s benefit to decrease salaries

24a. Subject _nominal Vj Object
24b. Subject_ sentential Vj Object

6 Conclusion and Future Work

In this paper, we argued that the lack of surface syntactic constraints of Chinese is the essential reason of the lower performances of the Chinese parsing trained on CTB than the English parsing trained on PTB. We gave examples to show that surface syntactic constraints of Chinese are less than English. We presented evidences to show that there exist strict deep syntactic constraints in Chinese sentences but CTB rules cannot effectively describe such constraints. We showed how to describe such deep syntactic constraints precisely based on SSG and how to develop a broad-coverage SSG-based Chinese grammar. The evaluating experiment was done and the results showed that the coverage of the Chinese grammar is 94.2%.

The CSSG rules analyze Chinese sentences deeper and more precisely than the CTB rules, so we will attempt to use it for Chinese parsing in the future.

References

Daniel M. Bikel and David Chiang. 2000. Two statistical parsing models applied to the Chinese Treebank. In Second workshop on Chinese language processing, volume 12, pages 1-6. Morristown, NJ, USA.

David Chiang and Daniel M. Bikel. 2002. Recovering latent information in treebanks. In Proceedings of the 19th international conference on Computational linguistics, volume 1, pages 1-7. Association for Computatical Linguistics.

Daniel M. Bikel. 2004. On the parmeter space of generative lexicalized statistical parsing models. Ph.D. thesis, Citeseer.

Eugene Charniak. 2000. Amaximum-entropy-inspired parser. In Proceedings of NAACL.

Liang Huang, Wenbin Jiang, and Qun Liu. 2009. Bilingually-constrained (monolingual) shift-reduce parsing. In Proceedings of the 2009 conference on Empirical Methods in Natural Language Processing, volume 3, pages 1222-1231. Association for Computational Linguistics.

Mengqiu Wang, Kenji Sagae, and Teruko Mitamura. 2006. A fast, accurate deterministic parser for Chinese. In Proceedings of the 21st International Conference on Computational Linguistics, pages 425-432. Association for Computational Linguistics.

M. Marcus, B. Santorini, M.A. Marcinkiewicz. Building a large annotated corpus of English: the Penn TreeBank. Computational linguistics. Vol 19, 1993.

Michael Collins. 2000. Discriminative reranking for natural language parsing In Proceedings of ICML, pages 175-182. Morgan Kaufmann, San Francisco, CA.

Nianwen Xue and Fei Xia. 2000. The bracketing Guidelines for the Penn Chinese Treebank.

Roger Levy and Christopher Manning. 2003. It is harder to parse Chinese, or the Chinese Treebank? In Annual Meeting of the Association for Computational Linguistics, volume 1, pages 439-446. Morristown, NJ, USA.

Xiangli Wang, Masahiro Miyazaki. 2007. *Chinese Syntactic Analysis Using Sentence Structure Grammar(in Japanese)*. Journal of Natural Language Processing. vol.14, No.2. April 2007

Xiangli Wang, Yusuke Miyao and Yuan Li. 2012a. *Chinese Grammatical resources based on Sentence Structure Grammar and its application on patent field (in Japanese)*. Proceeding of Japan Natural Language Processing. 2012.

Xiangli Wang, Terumasa Ehara and Yuan Li. 2012b. Parsing Simplified Chinese and Traditional Chinese with Sentence Structure Grammar. In proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing, Pages 179-187.

Yue Zhang and Stephen Clark. 2009. Transition-based parsing of the Chinese Treebank using a global discriminative model. In Proceedings of the 11th International Conference on Parsing Technologies, pages 162-171. Association for Computational Linguistics.