Parsing Simplified Chinese and Traditional Chinese with Sentence Structure Grammar

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

Terumasa Ehara
Yamanashi Eiwa College,
Yamanashi, Japan
eharate@{yamanashi-eiwa, y-eiwa}.ac.jp

Yuan Li
The University of Tokyo,
Tokyo, Japan
liyuan@is.s.u-tokyo.ac.jp

Abstract

We present a challenge to parse simplified Chinese and traditional Chinese with a same rule-based Chinese grammatical resource---Chinese Sentence Structure Grammar: CSSG, which was developed based on a new grammar formalism idea: Sentence Structure Grammar: SSG. We participate in the simplified Chinese parsing task and the traditional Chinese parsing task of CLP 2012 with a same rule-based chart parser implemented the CSSG. The experiments show that the CSSG that was developed for covering simplified Chinese constructions can also analyze most traditional Chinese constructions.

1 Introduction

Chinese divides into simplified Chinese that is used in the mainland of China and Singapore, and traditional Chinese that is used in Taiwan and Hang Kong. Some treebank resources like Penn Chinese Treebank: CTB, Peking University Treebank: PKU, and Tsinghua Chinese Treebank: TCT had been built for training simplified Chinese parser (Yu, et al. 2010) while Sinica Treebank was developed for parsing traditional Chinese (Chen et al., 1999). Limit to our knowledge, there are still not grammatical resources that analyze both simplified Chinese and traditional Chinese.

Recently, a rule-based Chinese grammatical resource --- Chinese Sentence Structure Grammar: CSSG had been developed based on the idea of Sentence Structure Grammar: SSG (Wang and Miyazaki, 2007; Wang et al., 2011, Wang et al., 2012). The CSSG was developed to cover most constructions that are listed in well-discussed simplified Chinese grammatical literatures (Zhu, 1982; Liu et al., 1996; Fan, 1998; Xue and Xia, 2000), and many phenomena that are not discussed in above literatures but very typical and used frequently by Chinese native speakers.

We assume that a rule-based grammatical resource should analyze both simplified Chinese and traditional Chinese if there are no obvious differences between their grammatical constructions. Aiming at verifying this assumptions, we participate in the simplified Chinese parsing task (task 3) and the traditional Chinese parsing task (task 4) of CLP 2012 with the same rule-based parser that was implemented the grammatical resource CSSG.

CSSG includes two parts of resources: the grammatical rules and a simplified Chinese morphological dictionary. We transfer the simplified Chinese characters of the dictionary to traditional Chinese characters for obtaining a traditional Chinese morphological dictionary. We parse the test data of task 3 and task 4 with the same CSSG rules but different morphological dictionaries (simplified or traditional Chinese characters). We convert CSSG parsing trees to TCT-style trees and Sinica-style trees for participating in the evaluations of the two tasks. The experiments show that the CSSG rules can parse both simplified Chinese and traditional Chinese, but

Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing, pages 179–187,
Tianjin, China, 20-21 DEC. 2012

the performance of the latter is lower than the former. We noticed that a few traditional Chinese constructions are different from simplified Chinese.

This paper is organized as bellow: in section 2, we introduce what is CSSG; in section 3, we compare CSSG with TCT and Sinica Treebank; in section 4, we analyze the experimental results of the two tasks; in the last section, we conclude our work.

2 Chinese Sentence Structure Grammar

Chinese Sentence Structure Grammar: CSSG is a rule-based Chinese grammatical resource that was developed based on the idea of Sentence Structure Grammar: SSG.

SSG is a new idea to formalize grammatical rules. Sentence Structure Grammar has 3 main ideas (Wang et al., 2011; Wang et al., 2012):

1) Treat the construction of a sentence as a whole, which consists of a predicate (or more) and its semantic-related constituents.
2) Classify predicate verbs according to their semantic properties.
3) Indicate the semantic relations between predicate and its semantic-related constituents directly on parsing tree.

Predicate	Ex.	Sematic roles
Vad	飞/fly	Agent, Direction
Vaod	扔/throw	Agent, Object, Direction
Vaol	放/put	Agent, Object, Location

Table 1: examples of the predicate classification of CSSG

SSG is a kind of context-free grammar, but it differs from Phrase Structure Grammar: PSG: 1) the latter describes a sentence with some context-free phrase rules, but the former treats a sentence as a whole sentential construction, which consists of a predicate (or more) and its semantically-related constituents; 2) the former classify predicate verbs according to their semantic properties. For instance, as shown in figure 1, "停/park" and "飞/fly" have different semantic properties. "停/park" is a kind of verb that needs an agent, an object and a location. In contrast, "飞/fly" is a kind of verb that needs an agent and a direction. As shown in table 1, predicate verbs can be classified according to their semantic properties; 3) the latter does only syntactic analysis while the former does syntactic analysis and semantic analysis simultaneously. The semantic role set of SSG should be designed based on the idea of the deep cases in Case Grammar, which a linguistic theory proposed by Fillmore (1968).

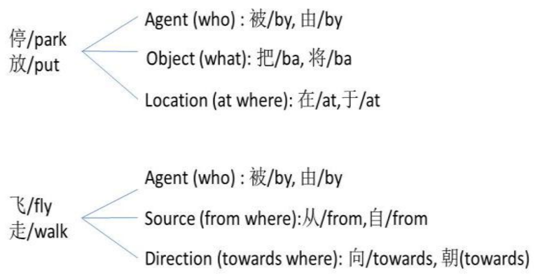

Figure 1: the semantic properties of verbs like "停/park" and "飞/fly"

For instance, a) is a passive construction. b) is the PSG rule set while c) is the SSG rule set to analyze a). Figure 2 and figure 3 show the SSG parsing tree and the PSG parsing tree of a). As shown in figure 2, the SSG parsing tree provide not only syntactic information like "np" and "sp" but sematic roles, like "Agent", "Object" and "Location", which indicate the semantic relations between the predicate and its semantic-related constituents. Syntactic parsing and sematic parsing can be done simultaneously with the formal grammatical framework SSG.

a. 车/car 被/by 约翰/John 停/park 在/at 停车场/car-park
The car is parked at the car-park by John

b. Rule1:s→ np vp
Rule2: vp→ pp vp
Rule3: vp→ v pp
Rule4: pp→p np
Rule5: np→ n
Rule6: sp→ sq

c. Rule1: s → Object bei Agent Vaol at Location
Rule2: Object→ np
Rule3: Agent→ np
Rule4: Location→ sp
Rule5: np→ n
Rule6: sp→ sq

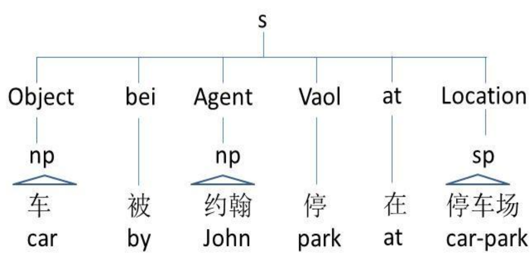

Figure 2: the SSG parsing tree of (a)

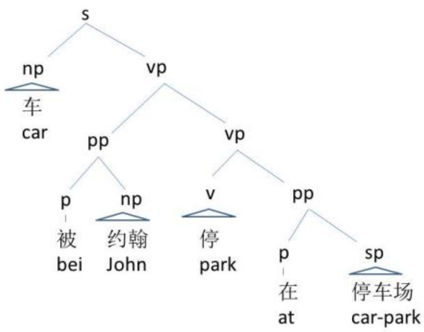

Figure 3: the PSG parsing tree of (a)

In CSSG, predicates are classified into 52 types according to their sematic properties. Table 1 shows some examples of the predicate classification. For instance, the type of verbs like “买/buy” or “拿/take” have same semantic property. They correspond to the same predicate-argument structure that is shown as figure 4. In CSSG, such semantic relations between a predicate and its arguments are showed on parsing rules directly. For instance, figure 5 shows the CSSG parsing tree of d): “买/buy” is the predicate, “他/he” is the agent case, “书店/bookshop” is the source case, “书/book” is the object case and “家/home” is the goal case. “把/ba” and “回/back” are treated as case-markers.

d. 他/he 从/from 书店/bookshop 把/ba 书/book 买/buy 回/back 家/home
He buys a book at the bookshop and takes it back home.

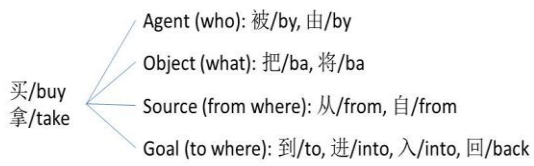

Figure 4: the semantic properties of the verbs like “买/buy” or “拿/take”

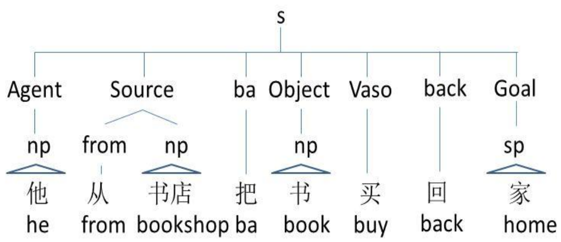

Figure 5: the CSSG parsing tree of (d)

CSSG includes two parts of resources: 8,511 grammatical rules and a morphological dictionary that contains 45,086 morphological entries.

The CSSG rules cover most constructions of simplified Chinese. Besides most constructions are listed in well-discussed simplified Chinese grammatical literatures (Zhu, 1982; Fan, 1998; Liu et al., 1996; Xue and Xia, 2000), the CSSG rules also cover many phenomena that were not discussed in above literatures but very typical and used frequently. For instance, e) is a ba-construction, f) is a bei-construction, g) is a topic construction, h) is not only a topic construction but a ba-construction and i) is not only a ba-construction but also a bei-construction. We observed many phenomena and found that there is a common feature in these different constructions, it is that one noun phrase “苹果 皮/skin of apples” is split into two parts, which have possessive relation each other but appear different syntactic positions in a sentence. Such constructions are called as “***apple-skin constructions***” in CSSG, and the possessive relation between the two split parts is indicated on the parsing tree. The CSSG rules analyze e), f), g), h) and i) as shown in figure 6, 7, 8, 9 and 10. “Object_of0” and “ Ojbect_of1” show the possessive relation between “苹果/apple” and “皮/skin”. Apple-skin constructions are used frequently by Chinese native speakers. We can make many sentences with them.

e. 约翰/John 把/ba 苹果/apple 削/peal 了/le 皮/skin
John pealed the apple’s skin

f. 苹果/apple 被/by 约翰/him 削/peal 了/le 皮/skin
The skin of apples was pealed by John

g. 苹果/apple 约翰/John 削/peal 了/le 皮/skin
The apple, John pealed its skin

h. 苹果/apple 约翰/John 把/ba 皮/skin 削/peal 了/le
The apple, John pealed its skin

i. 苹果/apple 被/by 约翰/John 把/ba 皮/skin 削/peal 了/le
The skin of apples was pealed by John

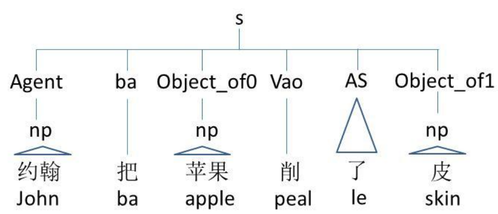

Figure 6: the CSSG parsing tree of (e)

The morphological dictionary of the CSSG includes two kinds of information: the morphology and its POS tag. Table 2 shows a small morphological dictionary for parsing a). The CSSG dictionary contains 45,086 simplified Chinese morphology entries. Table 3 shows the details of the dictionary. The word segmentation criteria and POS tag set of the CSSG were designed originally.

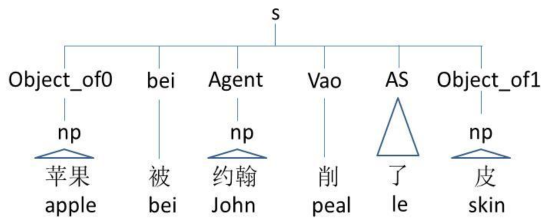

Figure 7: the CSSG parsing tree of (f)

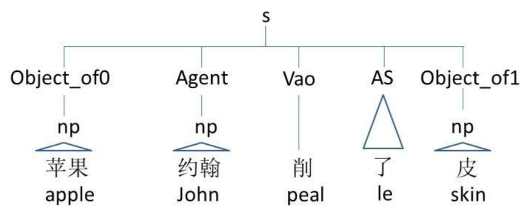

Figure 8: the CSSG parsing tree of (g)

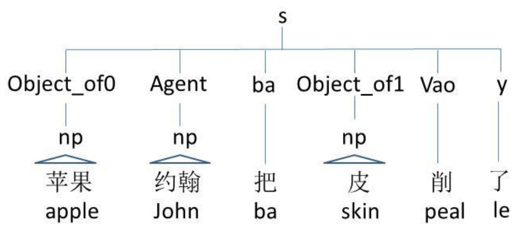

Figure 9: the CSSG parsing tree of (h)

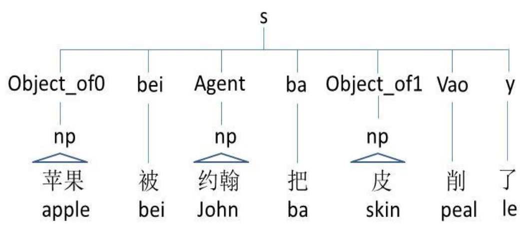

Figure 10: the CSSG parsing tree of (i)

Comparing with the existing Chinese treebanks, the design of the POS tag set of CSSG has some distinctive features. The major differences are: 1) verbs are classified according to their sematic properties; 2) some functional words are treated as a part of verbs in the existing treebanks are treated as Case-markers; 3) the localizers are divided into locative localizers and temporal localizers.

For instance, "买回/buy-back" is treated as one word in either TCT or CTB or Sinica Treebank, but in CSSG, as shown in figure 5, "买回/buy-back" is split into two words: "买/buy" and "回/back". "买/buy" is a predicate verb while "回/back" is a case-maker that marks a goal case.

Word	POS tag
车/car	n
约翰/John	n
停/park	Vaol
在/at	at
停车场/car-park	sq
被/by	bei

Table 2: a small dictionary for parsing (a)

part-of-speech	amount
verbs	6,878
nouns	26,191
adverbs	1,992
nominal verbs	5,028
temporal words	865
locative words	151
noun-modifier	2,439
measure words	446
pronouns	49
modal verbs	23
case markers	45
locative localizer	15
temporal localizer	17
others	947
total	45,086

Table 3: the details of the CSSG dictionary

j. 桌子/table 后/behind
Behind the table

k. 回/go-back 家/home 后/after
After going back home

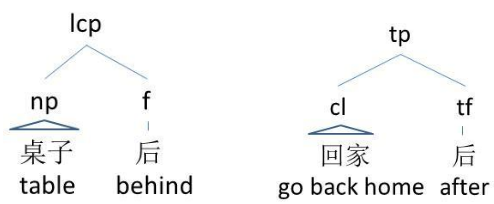

Figure 11: the CSSG parsing trees of (j) and (k)

In every existing Chinese treebank, the words like "前" and "后" are treated as localizers. However, either "前" or "后" contains two notions: a locative notion and a temporal notion. For instance, "后/behind" in j) shows a location while "后/after" in k) indicates a period of time. In CSSG, such words are divided into two kinds of POSs: locative localizers and temporal localizers. A locative localizer leads a locative phrase while a temporal localizer leads a temporal phrase (as shown in figure 11).

3 Comparison between TCT, Sinica Treebank and CSSG

3.1 Tsinghua Chinese Treebank and CSSG

Tsinghua Chinese Treebank: TCT (Zhou, 2004) is used as the training data for the simplified Chinese parsing task. TCT and CSSG are very different grammatical resources.

	CSSG	TCT
Formalism	SSG	PSG
Form	Grammatical rules	Treebank
Word segmentation criteria	Original	Original
POS tag set	Original	Original
Phrase tag set	Original	Original
Semantic role set	Original	none

Table 4: the differences between CSSG and TCT

Their main differences are: 1) they were developed based on different formal grammatical framework. As shown in figure 2 and 3, the former is based on Context-free Phrase Structure Grammar: PSG while the latter is based on another kind of Context-free grammar formalism idea---Sentence Structure Grammar: SSG. Since PSG parses sentences in syntactic level but SSG analyze sentences more deeply, CSSG provides both syntactic information and semantic roles while TCT shows only syntactic information. Figure 2 is a CSSG parsing tree of a) that represents both phrase information and sematic role information. Figure 3 is a TCT parsing tree that shows only syntactic information; 2) CSSG is a rule-based grammatical resource while TCT is a Treebank. The designers and developers of the treebanks are usually different people. The designers draw up the annotation scheme first, then the developers annotate parsing trees according to the annotation scheme and their own intuition; in contrast, the designer and the developer of CSSG is the same person who designed and developed the CSSG rules introspectively to cover most simplified Chinese constructions; 3) both of them define the word segmentation criteria and POS tag set originally. For instance, as shown in figure 12 and figure 13, TCT treats “来自/come-from” as one verb while CSSG treats “来自/come-from” as two words: “来/come” is a predicate verb and “自/from” is treated as a case-marker that mark a source case for describing semantic roles precisely; 4) they design the phrase tag set originally. As shown in figure 12 and 13, verb phrases appear in TCT while there are no verb phrases in CSSG; their definitions of prepositional phrase are different; as shown in figure 11: CSSG and 14: TCT, both j) and k) are treated as locative phrases in TCT while j) is treated as a locative phrase and k) is treated as a temporal phrase in CSSG. Table 4 shows the differences between TCT and CSSG briefly.

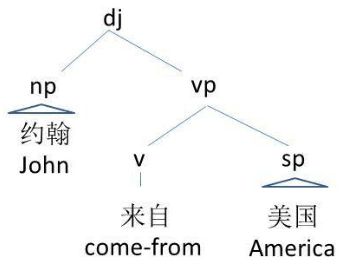

Figure 12: the TCT parsing tree of (l)

l. 约翰/John 来/come 自/from 美国/America
John comes from America

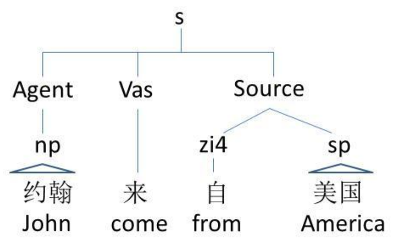

Figure 13: the CSSG parsing tree of (l)

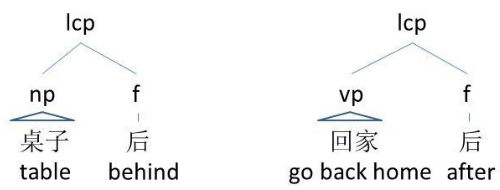

Figure 14: the TCT parsing trees of (j) and (k)

3.2 Sinica Treebank and CSSG

Sinica Treebank (Chen et al., 1999) is used as the training data for the traditional Chinese parsing task. CSSG are quite different from Sinica Treebank.

m. 那個/that 人/person 把/ba 老鼠/rat 帶/take 回/ back-to 茅屋/cottage
That man takes the rat back to the cottage

n. 約翰/John 從/from 房間/room 拿/take 出/out 一本/a 書/book
John takes a book out of the room

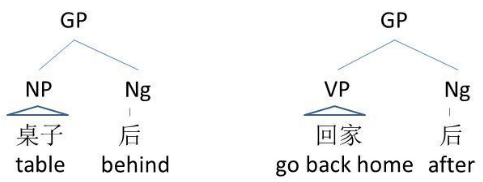

Figure 15: the TCT parsing trees of (j) and (k)

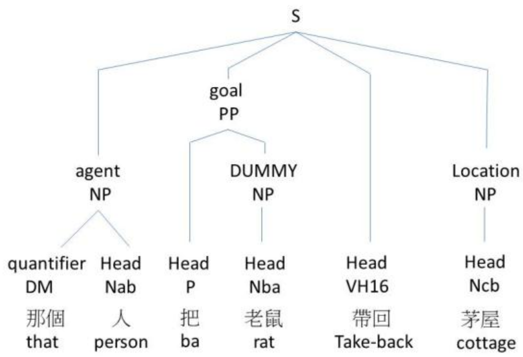

Figure 16: the Sinica parsing tree of (m)

They differ from each other in 6 respects: 1) Sinica Treebank consists of traditional Chinese parsing trees while CSSG is developed for covering simplified Chinese constructions; 2) the former is a rule-based grammatical resource while the latter is a Treebank; 3) both Sinica Treebank and CSSG represent syntactic and semantic information simultaneously, but their formal grammatical framework are different. Sinica Treebank is based on Information-based Case Grammar: ICG, which is a kind of unification-based formalism, and describe syntactic and semantic information in lexical entries (Chen and Huang, 1990); in contrast, CSSG is based on Sentence Structure Grammar: SSG, which is a kind of context-free grammar formalism that indicate both syntactic and semantic constraints in grammatical rules directly; 4) they define the word segmentation criteria and POS tag set originally. For instance, as figure 16 and 17 shown, "那個/that" is treated as one word in Sinica Treebank, but treated as two words in CSSG. "帶回/take-back" is one word in Sinica Treebank while it is split into a verb "帶/take" and a case-marker "回/back" that marks a goal case "茅屋/cottage" in CSSG; 5) they define the phrase tag set originally. For instance, the word "后" can lead not only a locative constituent like j) but a temporal constituent such as k). In Sinica Treebank, Both j) and k) are analyzed as a locative phrase (shown in figure 15); in contrast, the locative constituent is treated as a locative phrase while the temporal constituent is treat a temporal phrase in CSSG (shown in figure 10); 6) they define semantic role set originally. Their designs of the semantic role sets are very different. Figure 16 shows the Sinica-tree while figure 17 represents the CSSG tree of m). "老鼠/rat" is treated as a goal case and "茅屋/cottage" is analyzed as a location case in Sinica Treebank while "老鼠/rat" is regarded as an object case and "茅屋/cottage" is analyzed as a goal case in CSSG. As shown in figure 18 and 19, the source case "從/from 房間/room" in CSSG is treated as a location case in Sinica Treebank. Table 5 shows the differences between these two resources briefly.

	CSSG	Sinica Treebank
Character	Simplified	Traditional
Formalism	SSG	ICG
Form	Grammatical rules	Treebank
Word segmentation criteria	Original	Original
POS tag set	Original	Original
Phrase tag set	Original	Original
Semantic role set	Original	Original

Table 5: the differences between CSSG and Sinica Treebank

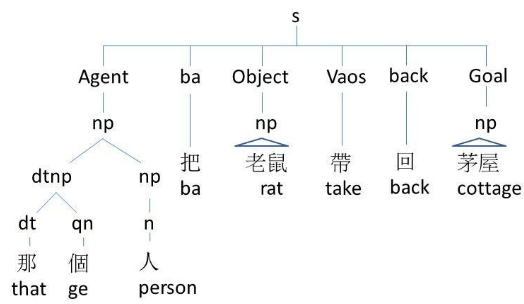

Figure 17: the CSSG parsing tree of (m)

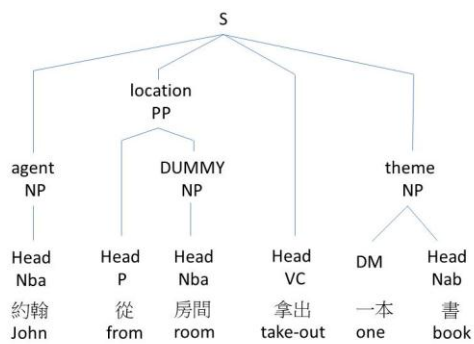

Figure 18: the Sinica parsing tree of (n)

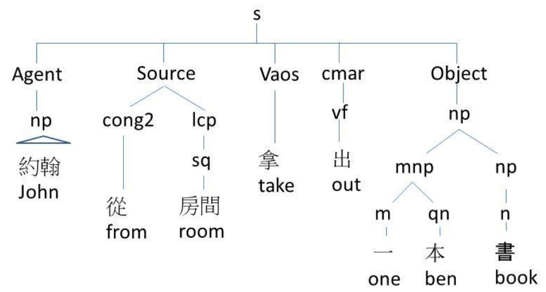

Figure 19: the CSSG parsing tree of (n)

4 Experimental Results

4.1 Experimental Setting

There are two parsing tasks in CLP2012: the simplified Chinese parsing task (task 3) and the traditional Chinese parsing task (task 4). Task 3 includes two subtasks: CCG parsing task and PSG parsing task while Task 4 includes two subtasks: sentence parsing task and semantic role labeling task. For each sub-task, there are two tracks: the closed track and the open track. Our tasks are all in the open tracks. We participate in the open tracks of the PSG parsing sub-task of task 3 and both the two sub-tasks of task 4.

CSSG includes the grammatical rules and a simplified Chinese morphological dictionary. For participating in both simplified Chinese parsing task and traditional Chinese parsing task, we transfer the simplified Chinese characters of the dictionary of CSSG to traditional Chinese characters to obtain a traditional Chinese morphological dictionary.

For instance, the simplified Chinese sentence a) can be transferred into a traditional Chinese sentence o). As shown in figure 2 and 20, a) and o) have the same construction. We can parse o) also with CSSG if there was a traditional Chinese morphological dictionary shown in table 6. We can transfer the small dictionary shown in table 1 to traditional Chinese characters to obtain the dictionary shown in table 6.

Word	POS tag
車/car	n
約翰/John	n
停/park	Vaol
在/at	at
停車場/car-park	sq
被/by	bei

Table 6: some samples of traditional CSSG dictionary

o. 車/car 被/by 約翰/John 停/park 在/at 停車場/car-park
The car is parked at the car-park by John

We parse simplified Chinese test data from task 3 with the parser implemented the grammatical rules and the simplified morphological dictionary while we parse the traditional Chinese test data from task 4 with the parser implemented the same grammatical rules and the traditional morphological dictionary. Since the scale of the dictionaries is not large enough, there are some unknown words for CSSG in both test data of task 3 and task 4. We add the unknown words to CSSG dictionaries before parsing.

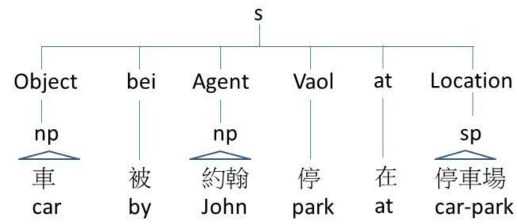

Figure 20: the CSSG parsing tree of (o)

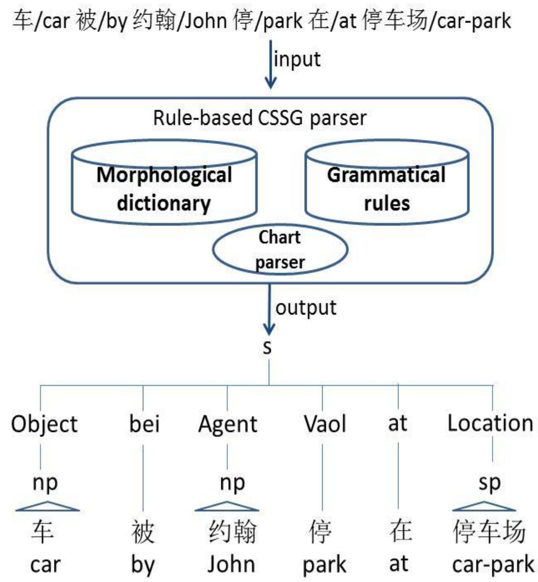

Figure 21: the input and output of (a) of the CSSG parser

As figure 21 shown: 1) the CSSG parser consists of three parts: the grammatical rules, a morphological dictionary and a chart parsing engine; 2) the input is a word-segmented sentence and the output is a CSSG parsing tree; 3) since there is not yet a postager based on CSSG, we have to parse all possible POS tag lists of a sentence with the CSSG parser.

After parsing the test data, we convert the CSSG parsing trees and make them are as similar as possible to TCT trees and Sinica-Treebank trees.

4.2 Evaluation Results

Table 7, 8 and 9 summarize the evaluation results of the simplified Chinese parse task 1: PSG parsing evaluation.

Table 7 shows the performance of the POS tag conversion from CSSG to TCT. Table 8 shows the results of the constituent boundary recognition. Table 9 represents the evaluation results of the parsing (both phrase boundaries and phrase labels recognition).

type	P	R	F1
nouns	74.4%	87.9%	80.6%
verbs	94.1%	94.1%	94.1%
others	62.7%	56.9%	59.7%
overall	71.3%	71.3%	71.3%

Table 7: the result for POS tag recognition

correct	gold	system	P	R	F1
85	92	158	53.8%	92.4%	68.0%

Table 8: the result for phrase boundary recognition

correct	gold	system	P	R	F1
85	92	158	42.4%	72.8%	53.6%

Table 9: the result for both phrase boundary and label recognition

Table 10 and 11 summarize the evaluation results of the two subtasks of the traditional Chinese parsing task. Table 10 presents the results of the parsing sub-task while table 11 shows the results of the semantic labeling sub-task.

Micro-averaging			Macro-averaging		
P	R	F1	P	R	F1
47.7%	40.1%	43.6%	53.6%	42.0%	47.1%

Table 10: the results of the parsing task

Micro-averaging			Macro-averaging		
P	R	F1	P	R	F1
20.4%	22.6%	21.4%	23.3%	24.2%	23.7%

Table 11: the results for the semantic labeling task

4.3 Discussion

As we anticipated, the evaluation results are lower than the real performance of the CSSG parser.

There are three reasons should be considered: 1) because of the large differences between the design of CSSG and the two gold data: Sinica Treebank and CSSG, it is impossible to convert some CSSG trees to TCT trees or Sinica-Treebank trees. For instance, k) is treated as a temporal phrase in CSSG, so it does not correspond to any phrase in TCT or Sinica Treebank; 2) there is much inaccuracy in tree-conversion works. As shown in table 7 and 8, the system phrase counts is 158, that is much more than the gold phrase counts 92 so that the recall scores (92.4% and 72.8%) are much higher than the precision scores (68.0% and 53.6%). We checked the evaluation data and found that we converted noun phrases of CSSG like p) to TCT format like q), which might be counted as two noun phrases; 3) As shown in figure 16, 17, 18 and 19, the design of the semantic role set of CSSG are very different from Sinica Treebank, so we can only convert a small number of sematic roles correctly.

p. (np (nnp (n 葡萄牙) (n 政府)))

q. (np (np (n 葡萄牙) (n 政府)))

As discussed above, the evaluation results do not reflect the real performance of the CSSG parser because of the large differences between CSSG and the two gold data. We expect that more neutral evaluation metrics would be drawn up for the open parsing task.

The experiments show that the evaluation results of the traditional Chinese parsing task are lower than the simplified Chinese parsing task. One of the possible reasons is that there are some differences between the constructions of simplified Chinese and traditional Chinese. We noticed that a few traditional Chinese constructions differ from simplified Chinese. For example, in traditional Chinese sentence r), "食/food" is the direct object that appears at the left side of the indirect object "企鵝寶貝/penguin-baby". We had ever asked some Chinese native speakers whether they think the construction like r) is grammatical. Only one speaker who comes from Hang Kong thinks r) is a grammatical sentence while other speakers who come from the mainland of China think such constructions are ungrammatical. Therefore simplified Chinese sentence s) is an ungrammatical sentence. For Simplified Chinese native speakers, a function word "给/to" should be used to lead an indirect object, like t) and u), or the indirect object appears at the left side of the direct object, such as v). The CSSG rules cover the constructions of t), u) and v) but not cover the constructions of r) and s).

r. 工作人員/worker 每天/every-day 仍會/yet 餵/ feed 食/food 企鵝寶貝/penguin-baby
The worker feeds foods to penguin babies everyday

s. *工作人员/worker 每天/every-day 仍会/yet 喂/ feed 食/food 企鹅宝贝/penguin-baby
The worker feeds foods to penguin babies everyday

t. 工作人员/worker 每天/every-day 仍会/yet 喂/ feed 食/food 给/to 企鹅宝贝/penguin-baby
The worker feeds foods to penguin babies everyday

u. 工作人员/worker 每天/every-day 仍会/yet 给/to 企鹅宝贝/penguin-baby 喂/ feed 食/food
The worker feeds foods to penguin babies everyday

v. 工作人员/worker 每天/every-day 仍会/yet 喂/ feed 企鹅宝贝/penguin-baby 食/food
The worker feeds foods to penguin babies everyday

5 Conclusion and Future work

In this paper, we introduced a broad-coverage rule-based Chinese grammatical resource CSSG, which was developed based on a new grammar formalism idea: Sentence Structure Grammar; we compared briefly CSSG with a simplified Chinese Treebank TCT and a traditional Chinese resource Sinica Treebank; we also introduced our participation of CIPS-SIGHAN-2012 parsing task. We use a same rule-based chart parser implemented CSSG to participate in both simplified Chinese parsing task and traditional Chinese parsing task. The experiment shows that the rule-based grammatical resource CSSG that was developed for covering simplified Chinese constructions can also parse traditional Chinese sentences with a lower performance.

Since the CSSG provide rich information, it is possible to improve the precision and the recall of the evaluation task by optimizing the tree-conversion programs.

We prepare to open this resource to researchers who have an interest in it in the resent future.

References

Chen Feng-Yi, Pi-Fang Tsai, Keh-Jiann Chen, Chu-Ren Hunag. 1999. *The Construction of Sinica Treebank*. Computational Linguistics and Chinese Language Processing, vol. 4, No. 2. pp.87-104.

Chen Keh-jiann and Chu-Ren Huang. 1990. *Information-based Case Grammar*. Proceedings of the 13th Conference on Computational Linguistics, Volume 2, pages 54-59.

Fan Xiao. 1998. *The types of Chinese Sentences (In Chinese)*. Shanxi Shuhai Press.

Fillmore, Charles J. (1968). *The Case for Case*. In Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88.

Liu Yuehua, Wenyu Pan and Wei Gu. 2001. *Practical Modern Chinese Grammar (In Chinese)*. Beijing: Commercial Press.

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

Wang Xiangli, Takuya Matsuzaki, Yusuke Miyao, Kun Yu, Yuan Li and Junichi Tsujii. 2011. *Comparison between formal grammatical frameworks for Treebanking (In Japanese)*. Proceeding of Japan Natural Language Processing. 2011.

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

Xue Nianwen and Fei Xia. 2000. *The bracketing Guidelines for the Penn Chinese Treebank Project*. Technical Report IRCS 00-08, University of Pennsylvania.

Yu Kun, Yusuke Miyao, Takuya Matsuzaki, Xiangli Wang, Yaozhong Zhang, Kiyotaka Uchimoto, Junichi Tsujii. *Comparison of Chinese Treebanks for Corpus-oriented HPSG Grammar Development.* Journal of Natural Language Processing (Special Issue on Empirical Methods for Asian Language Processing). April 2010.

Zhu Dexi. 1982. *Lecture Notes on Grammar (In Chinese)*. Beijing: Commercial Press.

Zhou Qiang. 2004. *Annotation Scheme for Chinese Treebank (in Chinese)*. Journal of Chinese Information Processing, 18(4): 1-8.