

Analysis of the Difficulties in Chinese Deep Parsing

Kun Yu¹ Yusuke Miyao² Takuya Matsuzaki¹ Xiangli Wang³ Junichi Tsujii⁴

1. The University of Tokyo, Tokyo, Japan

{kunyu, matuzaki}@is.s.u-tokyo.ac.jp

2. National Institute of Informatics, Tokyo, Japan

yusuke@nii.ac.jp

3. Japan Patent Information Organization, Tokyo, Japan

xiangli_wang@japio.or.jp

4. Microsoft Research Asia, Beijing, P.R.China

jtsujii@microsoft.com

Abstract

This paper discusses the difficulties in Chinese deep parsing, by comparing the accuracy of a Chinese HPSG parser to the accuracy of an English HPSG parser and the commonly used Chinese syntactic parsers. Analysis reveals that deep parsing for Chinese is more challenging than for English, due to the shortage of syntactic constraints of Chinese verbs, the widespread *pro-drop*, and the large distribution of ambiguous constructions. Moreover, the inherent ambiguities caused by verbal coordination and relative clauses make semantic analysis of Chinese more difficult than the syntactic analysis of Chinese.

1 Introduction

Syntactic parsing provides only the syntactic structure of text, while deep parsing offers richer information, such as the semantic roles. With the advancement of research in natural language processing, this rich information has become important for many applications, including statistical machine translation, information extraction, and question answering.

Performing semantic role labeling (Marquez et al., 2009) with shallow parsing is one way to fulfill deep parsing. Another alternative to semantic role labeling is to perform deep parsing based on lexicalized grammar theories, such as Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994), Lexical Functional Grammar (LFG) (Dalrymple et al., 1995), Combinatory Categorical Grammar (CCG) (Steedman, 2000), and Lexicalized Tree Adjoining Grammar (LTAG) (O'Donovan et al., 2005).

Many research projects have been done successfully in this way, such as is the case in parsing English with HPSG (Miyao and Tsujii, 2008; Matsuzaki et al., 2007), CCG (Clark and Curran, 2004), and LFG (Kaplan et al., 2004).

However, obtaining the deep analysis of Chinese has proven to be more difficult. We evaluated an existing HPSG parser, which has been used successfully for English deep parsing (Miyao and Tsujii, 2008), on the Chinese HPSG Treebank constructed by Yu et al. (2010). The results indicated that compared to English, this parser obtained a 12.97% decrease in semantic F1-score on Chinese deep parsing.

Therefore, this paper focuses on investigating the difficulties in Chinese deep parsing, by comparing the parsing results of this HPSG parser on both Chinese and English, with the parsing results from commonly used Chinese syntactic parsers. This is the first time that the difficulties in Chinese deep parsing were analyzed; the resulting analysis provides insight into future research for Chinese deep parsing.

2 Linguistic Properties of Chinese

As discussed in Guo (2009), Chinese has little inflectional morphology, compared with Indo-European languages. There is no tense, case, and number marker in Chinese, and in sequence, there are fewer syntactic constraints; such as the case with the agreement in English. Therefore, in Chinese, word order plays an important role in determining the sentence meaning.

吃/eat 过 苹果/apple 了。
(*Somebody*) has eaten the apple.)

(a) A Chinese sentence with subject *pro-drop*

吃/eat 过了。
 ((Somebody) has eaten (something).)

(b) A Chinese sentence with both subject and object *pro-drop*

Figure 1: Examples of *pro-drop* in Chinese

The other significant linguistic property in Chinese is the frequent *pro-drop* phenomena. For example, Levy and Manning (2003) showed that unlike English, the subject *pro-drop* (the null realization of uncontrolled pronominal subjects) is widespread in Chinese; this is exemplified in Figure 1 (a). Huang (1989) further provided a detailed analysis to show that subjects as well as objects may drop from finite Chinese sentences (as shown in Figure 1 (b)).

3 Chinese Deep Parsing based on HPSG

3.1 Parsing Model

In this paper, we used an HPSG parser - Enju¹, which was successfully applied in English deep parsing, to obtain the deep analysis of Chinese. This HPSG parser uses the feature forest model proposed by Miyao and Tsujii (2008), which is a maximum entropy model that is defined over feature forests, as a parsing disambiguation model. The feature forest model provides a solution to the problem of probabilistic modeling of complex data structures. Moreover, in order to reduce the search space and further increase the parsing efficiency, in this parser, a supertagger (Matsuzaki et al. 2007) is applied before parsing. This supertagger provides the maybe-parsable supertag (i.e. lexical template) sequences to the parser.

In short, in the HPSG parser, the probability, $p(t|w)$, of producing a parse tree t for a given sentence w is defined by Equation 1. Here, Z_w is a normalization constraint; $p(l|w)$ is a maxent supertagging model in which l is the supertag sequence for sentence w ; $f_i(t, l, w)$ is a feature function that represents the characteristics of t , l , and w ; and λ_i is its weight. When performing Chinese HPSG parsing, the feature functions (i.e. $f_i(t, l, w)$) were borrowed from the English parser without any change, but the weights (i.e. λ_i) were tuned by using the development data.

$$p(t|w) = \frac{1}{Z_w} p(l|w) \exp\left(\sum_i \lambda_i f_i(t, l, w)\right) \quad (1)$$

¹ <http://www-tsujii.is.s.u-tokyo.ac.jp/enju/index.html>

The parsing procedure of this HPSG parser can be explained in the following way:

Given a segmented and pos-tagged input sentence,

(1) the supertagger offers all the maybe-parsable supertag (i.e. lexical template) sequences with scores to the parser;

(2) the feature forest model applies beam threshold on the scored supertag sequences, and then obtains a well-formed HPSG parse tree.

Figure 2 shows a supertag sequence provided by the supertagger for a Chinese sentence, in which the supertag of the word ‘写(wrote)’ indicates a lexical template for the transitive verb with an extracted object. Figure 3 illustrates the HPSG parse tree output from the parser with this supertag sequence.

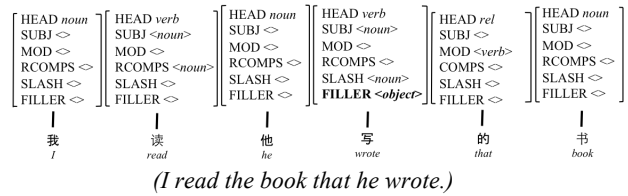


Figure 2: A supertag sequence provided by the supertagger

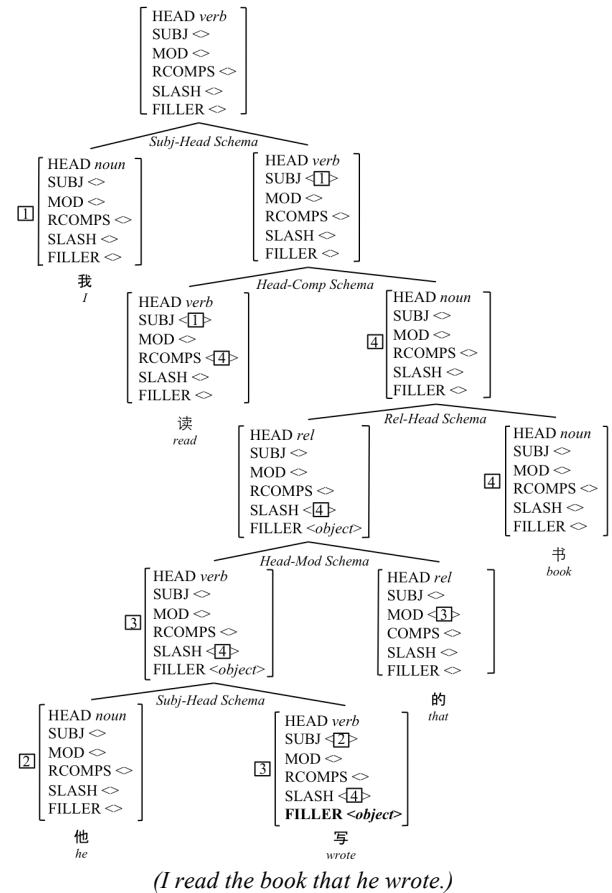


Figure 3: The HPSG tree created from Figure 2

3.2 Training Data

In order to apply the HPSG parser to Chinese deep parsing, we used the Chinese HPSG Treebank developed by Yu et al. (2010) to train the parser.

This Chinese HPSG Treebank is based on the Chinese HPSG grammar designed in (Yu et al., 2010). 25,724 (95.66%) trees in the Penn Chinese Treebank 6.0 were successfully converted into HPSG trees, with 97.24% accuracy (Yu et al., 2010). For the details concerning the construction phase, please refer to (Yu et al., 2010).

From the syntactic point-of-view, in addition to the phrase structure of the Penn Chinese Treebank, this HPSG Treebank records the syntactic dependency relations, which are identified with the head rules similar to the head rules provided by Yuan Ding².

This treebank uses 51 types of predicate-argument dependencies to represent the semantic structures among 13 classes of words. A predicate-argument dependency is defined as $\langle w_p, w_a, r, l \rangle$, where w_p is the head word of the predicate and w_a is the head word of the argument. r is the type of predicate-argument dependency between w_p and w_a . l is the argument label, such as *ARG1* and *ARG2*.

3.3 Experimental Setting

By using the Chinese HPSG Treebank described above, we re-trained the feature forest model and the supertagger, and built a Chinese HPSG parser. The treebank was split into development, testing, and training data sets, following the recommendation from the authors of the Penn Chinese Treebank. The training data was used to train the HPSG parser, and the testing data was used for parsing evaluation; the development data was used for parameter tuning. Table 1 shows the statistics that resulted from the different data sets.

Data Set	# Total Tree	# Success Tree	# Word	# Template
Train	22,224	21,186	557,447	2,185
Test	2,635	2,530	71,921	863
Dev	2,067	2,042	56,736	783

Table 1: Statistics for the Chinese HPSG Treebank

In the experiments performed with for the HPSG parser, the gold-standard word boundaries and POS tags were supplied.

² http://w3.msi.vxu.se/~nivre/research/chn_headrules.txt

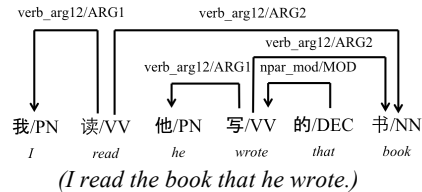


Figure 4: A predicate-argument dependency parse tree output by the Chinese HPSG parser

The Chinese HPSG parser offers predicate-argument dependencies as the output of semantic parsing. Figure 4 illustrates a parse tree with a predicate-argument dependency that has been built by the Chinese HPSG parser, in which the label of each dependency is the combination of r and l in a predicate-argument dependency $\langle w_p, w_a, r, l \rangle$. As an example, the predicate-argument dependencies of the verb ‘写(writes)’ shown in Figure 4 indicates that the verb is a transitive verb (*verb_arg12*), and has a subject (*ARG1*) ‘他(he)’, and an object (*ARG2*) ‘书(book)’.

Therefore, we evaluated the performance of the Chinese HPSG parser on semantic parsing by analyzing the accuracy of the predicate-argument dependencies. Six evaluation metrics used by Miyao and Tsujii (2008) were selected for the evaluation. *LP* and *LR* refer to the labeled precision and recall of the predicate-argument dependencies, while *UP* and *UR* refer to the unlabeled precision and recall, respectively. *Sem.F1* is the semantic F1-score calculated based on *LP* and *LR*. *Sent.acc.* is the accuracy of the sentences with the correct predicate-argument dependencies.

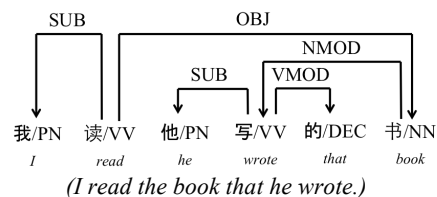


Figure 5: A syntactic dependency parse tree corresponding to Figure 4

Besides of semantic analysis, the Chinese HPSG parser also provides the syntactic head for each branch in an HSPG parse tree and the schemas used to construct the branch, which can be used to extract the labeled syntactic dependency as the output of syntactic parsing. In order to evaluate the syntactic analysis of the Chinese HPSG parser, we used the similar dependency labels as the CoNLL dependency labels (Nivre et al., 2007 (b)). Figure 5 shows the labeled syntactic dependency tree output by the parser, in which the label *SUB* and *OBJ* refer to the subject

and object, respectively. The common metrics used in CoNLL-2007 shared task (Nivre et al., 2007 (b)) were applied in the evaluation of the syntactic parsing. These metrics include the labeled attachment score (*LAS*), unlabeled attachment score (*UAS*), and the complete sentence accuracy (*COMP*) with labeled dependency.

3.4 Evaluation Results

The accuracy of both syntactic parsing and semantic parsing of the Chinese HPSG parser was 83.75% *LAS* and 77.55% *Sem.F1*, and is listed in Table 2 and Table 3.

To compare the performance of the Chinese HPSG parser on syntactic parsing with other related works, we evaluated two commonly used syntactic dependency parsers: MaltParser (Nivre et al., 2007 (a)) and MstParser (McDonald et al., 2006); the same syntactic dependency converted from the Chinese HPSG Treebank was used. In this experiment, the MaltParser and MstParser used both the gold-standard word boundaries and gold-standard POS tags, like the HPSG parser. Table 2 displays the results. The Chinese HPSG parser achieved a comparable accuracy to the MaltParser and the MstParser with 1st order features, but the Chinese HPSG parser’s accuracy was slightly lower than the accuracy of the MstParser with 2nd order features.

Parser	<i>LAS</i>	<i>UAS</i>	<i>COMP</i>
Chinese HPSG	83.75%	85.57%	29.67%
Malt	83.74%	84.17%	29.01%
MST (1 st order)	84.75%	85.22%	25.99%
MST (2 nd order)	86.44%	86.95%	30.54%

Table 2: Accuracy of syntactic parsing

<i>LP</i>	<i>LR</i>	<i>UP</i>	<i>UR</i>
77.14%	77.97%	81.82%	82.70%
<i>Sem.F1</i>		<i>Sentence acc.</i>	
77.55%		23.84%	

Table 3: Accuracy of semantic parsing by the Chinese HPSG parser

Parser	<i>Sem.F1</i>
(Bjorkelund et al., 2009)	78.60%
(Meza-Ruiz and Riedel, 2009)	77.73%
(Zhao et al., 2009)	77.72%

Table 4: Accuracy of the top three systems in CoNLL-2009 Shared Task on Chinese Data

Since there has been no previous work conducted on the same Chinese HPSG formalism as used in the HPSG parser, comparing our semantic parsing results against the results of the exist-

ing approaches would not be accurate. However, a closely related work on joint syntactic and semantic parsing was done in the CoNLL-2009 shared task (Hajic et al., 2009). In this shared task, the Penn Chinese Treebank and the Chinese Proposition Bank (Xue and Palmer, 2009) were merged to serve as the training and testing data, and a semantic labeled F1-score (*Sem.F1*) was applied to evaluate the performance of semantic role labeling (Hajic et al., 2009). While the CoNLL-2009 shared task only applied gold-standard word boundaries, our experiment used both gold-standard word boundaries and gold-standard POS tags.

Table 4 lists the performance of the top three systems on the closed challenge for Chinese in the CoNLL-2009 shared task. Unfortunately, we cannot compare the result of the Chinese HPSG parser to the results of the top three systems in the CoNLL-2009 shared task, because of the different experimental settings. However, all the top systems in the shared task performed semantic role labeling after the syntactic parsing from the state-of-the-art parsers took place, whereas in our experiment, the Chinese HPSG parser applied a joint model that performed syntactic parsing and semantic parsing at the same time.

4 Discussion Concerning the Difficulties in Chinese Deep Parsing

4.1 Chinese Deep Parsing vs. English Deep Parsing

The HPSG parser that we used for Chinese deep parsing was also applied for English deep parsing (Miyao and Tsujii, 2008). Thus, we first compared the performance of the HPSG parser on parsing Chinese and English. In this experiment, we applied the same supertagging model with the same definition of supertags and feature sets, and the same parsing disambiguation model with the same feature sets, to the two treebanks.

To parse English, we used the English HPSG Treebank, which has been developed by Miyao et al. (2006), to train and evaluate the parser. The design of this treebank basically followed the definition in (Pollard and Sag, 1994). The HPSG trees converted from Sections 02-21 (39,832 sentences) of the Penn Treebank were used for training. The HPSG trees transformed from Section 23 (2,416 sentences) of the Penn Treebank were used for evaluation, and the HPSG trees converted from Section 22 (2,067 sentences) were used to tune parameters.

Parser	English	Chinese
HPSG	90.52%	77.55% (-12.97%)
HPSG + gold supertag	95.66%	92.52% (-3.14%)

Table 5: *Sem.F1* of the HPSG parser on both English and Chinese, with different models

We evaluated two different parsers in this experiment: the HPSG parser introduced in Section 3, and the HPSG parser with the gold-standard supertag sequence as input. Table 5 lists the evaluation results for both English and Chinese data. The results indicate that compared to English, the HPSG parser obtained 12.97% decrease in *Sem.F1* when parsing Chinese. Furthermore, this result shows that when given the exact supertag sequence of an input sentence, the HPSG parser still achieved a lower (3.14%) accuracy on Chinese than on English.

Data	# Ave. Parses	# Ave. Words	# Ave. Verbs	Sentence Distribution (#Verb>3)
Eng. Dev.	10,988.74	23.20	2.97	35.88%
Chi. Dev.	37,200,740.79	26.37	4.66	59.36%

Table 6: Average number of parses, words, and verbs per sentence in the English and Chinese development data

Training data deficiency may account for the low accuracy when parsing Chinese. However, the learning curve shown in (Miyao and Tsujii, 2008) indicates that even with half of the size of the full training data (i.e. 24,000 sentences), the HPSG parser obtained similar accuracy values on English deep parsing. Furthermore, we counted the average number of parses per sentence when given the exact supertag sequence for both Chinese and English on the development data. The numbers (as listed in Table 6) indicate that the parsing disambiguation is more difficult for Chinese than for English, because Chinese sentences have much more parses averagely than English sentences given the exact supertag sequence. A possible reason for the large average number of parses in Chinese is that Chinese sentences contain more verbs than English (as shown in Table 6). Due to the shortage of syntactic constraints of Chinese verbs, such as the agreement in English, it is easier for Chinese sentences with verbs to create ambiguous parses than for English.

Moreover, the comparison of the overall supertagging accuracy on Chinese and English (as

shown in the left-most column in Figure 6) reveals that besides of the difficulty in Chinese parsing disambiguation, Chinese supertagging is also more difficult than for English. Following displays the possible reasons.

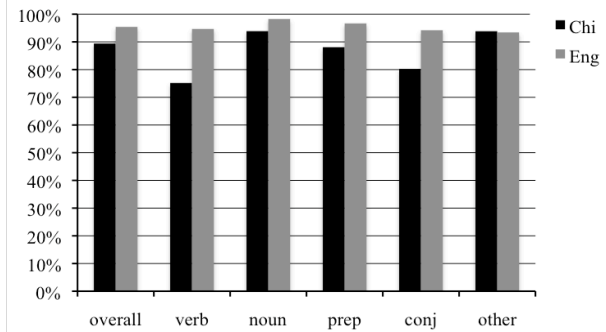


Figure 6: Supertagging accuracy on both English and Chinese testing data

(1) In comparison with English words, Chinese words have a much larger averaged number of supertags, especially for verbs.

Table 7 lists the total number of supertags and the average number of supertags per word in both the English HPSG Treebank and the Chinese HPSG Treebank. These numbers reveal that with the same granularity of supertags, although the total number of supertags is similar for both English and Chinese, the Chinese words have almost twice the average number of supertags than English words have. This difference makes it difficult for the supertagger to assign correct supertags for Chinese sentences.

Treebank	# Total Supertag	# Ave. Supertag for all words	# Ave. Supertag for verb
English	1,368	12.46	27.61
Chinese	1,279	21.57	87.82

Table 7: Statistics of the supertags in the English and Chinese HPSG Treebank

In addition, the analysis indicates that compared to other word types, the supertags of verbs in Chinese have more variations than verbs in English. As shown in Table 7, in the English HPSG Treebank, a verb has an average of 27.61 different supertags. By contrast, in the Chinese HPSG Treebank, a verb has an average of 87.82 different supertags. Table 8 lists the main reasons for the various verb supertags in Chinese and the sentence percentage with corresponding phenomena in the Chinese HPSG Treebank, of which the widespread subject *pro-drop* is the most predominant. The restrictions of the modifier and topic in the supertag definition also

brings about a large variation to the verb supertags. Changing the granularity of supertags of Chinese verbs is a possible way to solve this problem. Experimental results showed that by removing the restrictions of modifiee and topic in the definition of verb supertags, the *Sem.F1* could be improved by 0.3%.

Reason	Percentage
Subject <i>pro-drop</i>	34.75%
With/without modifiee	23.19%
With/without Topic	10.51%
Auxiliary verb	2.19%
Non-local dependency	0.30%

Table 8: Distribution of the main reasons for various verb supertags in Chinese

(2) *The ambiguous constructions in supertagging have a larger distribution in the Chinese HPSG Treebank than in the English HPSG Treebank.*

The supertagger’s performance on different types of words, as shown in Figure 6, implicates that compared to English, Chinese verbs obtained the largest decrease in the accuracy of supertagging; 21.23% of the errors were related to the relative clause. Figure 6 also shows that in addition to verbs, coordination conjunctions decreased the accuracy of supertagging in Chinese.

However, there is not much difference in the supertagging ambiguity of coordination and relative clause in the two languages. For example, for both Chinese and English, in the supertagging of a verb in the relative clause, there is ambiguity as to whether assigning extracted predicate-argument dependency to this verb; the supertagging of a comma between verb phrases, has ambiguity in whether this comma will be treated as a coordination conjunction. Therefore, we further calculated the percentage of the sentences, including the verbal coordination with a comma conjunction and the relative clause in the two treebanks.

Treebank	Relative clause	Verbal Coordination with comma conj
English	14.31%	9.31%
Chinese	33.26%	52.95%

Table 9: Distribution of constructions in the English and Chinese HPSG Treebank

The statistics data is shown in Table 9. It reveals that in the Chinese HPSG Treebank, there are much more relative clauses than in the English HPSG Treebank. Moreover, the proportion

of verbal coordination with comma conjunction in Chinese was also much larger than the proportion in English. Therefore, although the supertagging ambiguities of verbal coordination and relative clauses are similar for the two languages, the large distribution of these constructions increased the difficulty of Chinese supertagging.

4.2 Chinese Semantic Parsing vs. Chinese Syntactic Parsing

In comparing the accuracy of both the semantic parsing and syntactic parsing, as shown in Table 2 and Table 3, it is clear that although the performance on the syntactic analysis of the parser still has room for further improvement, the accuracy of predicate-argument dependencies was significantly lower than the accuracy of syntactic dependencies. Therefore, in this section, we focus on this gap by comparing the syntactic and semantic parsing results from the Chinese HPSG parser.

Error	# Occur
Subject of transitive verb	84
Left conjunct in coordination	84
Modifiee of punctuation	70
Root of sentence	51
Object of transitive verb	49
Right conjunct in coordination	46
Modifiee of noun	43
Modifiee of adverb	41
Subj. of intransitive verb	31
Missed object of transitive verb	28

Table 10: Occurrence of top 10 frequently occurring errors

We chose 93 sentences from the development data, which obtained a higher accuracy on syntactic parsing (i.e. with more than 85% *LAS*) and lower accuracy on semantic parsing (i.e. with less than 75% *Sem.F1*); the detailed errors were analyzed. The top 10 frequently occurring errors with their occurrence were documented in Table 10. The table indicates that there are two main difficulties in Chinese semantic parsing, in comparison to syntactic parsing.

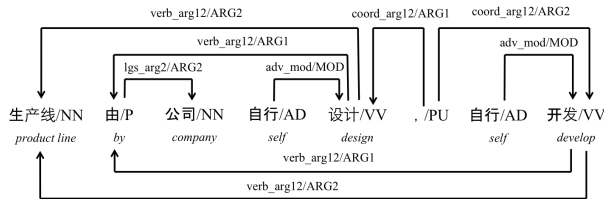
▪ Difficulty in Analyzing the Semantics of Parallel Verb Phrases

As indicated in Table 10, the top nine errors were attachment errors; for a predicate-argument dependency $\langle w_p, w_a, r, l \rangle$, only w_a is incorrect. There were 499 incorrect predicate-argument dependencies with the top nine errors. Of them,

59.12% of the errors were related to the semantic analysis of parallel verb phrases.

When two verb phrases are parallel, there are two possible semantic analyses for them:

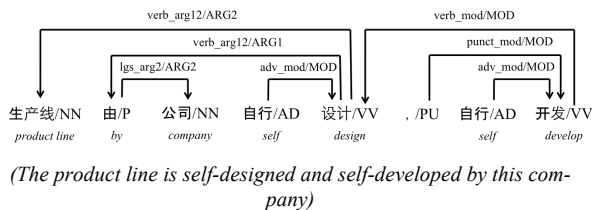
(1) The two verb phrases are treated as coordination, and consequently share the same subject. Figure 7 shows the predicate-argument dependency tree created by this analysis.



(The product line is self-designed and self-developed by this company)

Figure 7: The predicate-argument dependency tree when treating parallel VPs as coordination

(2) Treating the second verb phrase as a modifier of the first verb phrase. Figure 8 shows the corresponding predicate-argument dependency tree, in which the dependency *verb_arg12/ARG1* and *verb_arg12/ARG2* for verb ‘开发(develop)’ are missed.



(The product line is self-designed and self-developed by this company)

Figure 8: The predicate-argument dependency tree when treating parallel VPs as modification

However, the above such ambiguity in semantic analysis does not exist in the syntactic analysis of this type of construction. For example, no matter which type of semantic analysis the parser chooses, the syntactic dependency trees for the sentences shown in Figure 7 and Figure 8 are the same (as shown in Figure 9).

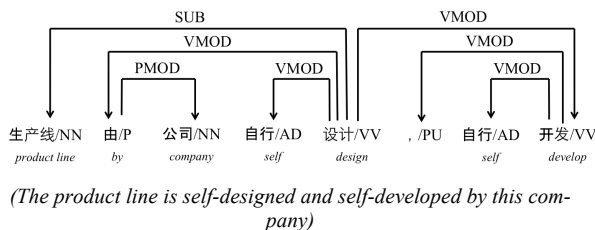


Figure 9: The syntactic dependency tree corresponding to Figure 7 and Figure 8

▪ **Difficulty in Analyzing the Semantics of Relative Clause**

The tenth error shown in Table 10 was a type of relation error, in which the parser failed to find the object for a transitive verb. The error analysis shows that among the 28 incorrect predicate-argument dependencies with this type of error, 71.43% of the incorrect predicate-argument dependencies were from the incorrect semantic analysis of the relative clause.

There are two possible ways to analyze the semantics of a relative clause. The first way is to analyze the extracted noun in a relative clause as a moved argument of the predicate. The second way is to treat the relative clause as an apposition of the following noun, such that the extracted noun has no semantic relation with the predicate. For example, among the relative clauses in the Chinese HPSG Treebank, about 81% of the clauses were analyzed in the first way, and the remaining 19% were analyzed in the second way.

In reference to the relative clauses shown below, for the relative clause ‘写书的人(the person who wrote the book)’, the semantics should be created by the first analysis (as shown in Figure 10), in which there is a predicate-argument dependency *verb_arg12/ARG1* between ‘写(wrote)’ and ‘人(person)’. However, for another relative clause ‘写书的原因 (the reason that someone wrote the book)’, the clause should be analyzed as an apposition (as shown in Figure 11). This is because the head noun ‘原因(reason)’ has no predicate-argument relation with the verb ‘写(wrote)’.

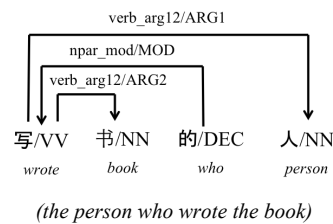


Figure 10: The predicate-argument dependency tree when analyzing a relative clause with extracted argument

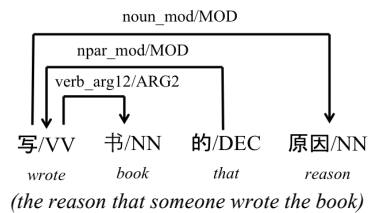


Figure 11: The predicate-argument dependency tree when treating a relative clause as an apposition

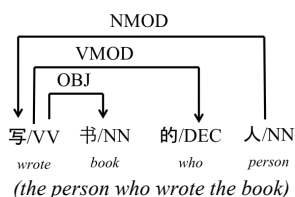


Figure 12: The syntactic dependency tree corresponding to Figure 10

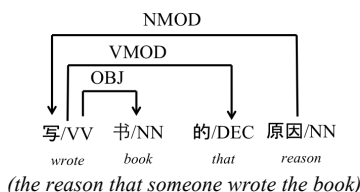


Figure 13: The syntactic dependency tree corresponding to Figure 11

However, since the syntactic analysis does not consider predicate-argument dependencies, such an ambiguity in semantic parsing does not exist in syntactic parsing. For instance, for both the two semantic analyses listed in Figure 10 and Figure 11, the syntactic dependencies are similar, as shown in Figure 12 and Figure 13.

5 Related Works

One related work was done by Levy and Manning (2003) on analyzing the difficulties in Chinese PCFG parsing. In this work, the authors applied a factored-model statistical parser on both the Penn Treebank (Marcus et al., 1994) and the Penn Chinese Treebank (Xue et al., 2005), and investigated the major sources of syntactic parsing errors and the corresponding causes in the two treebanks. The authors found that among the major error types in Chinese PCFG parsing, the coordination scope errors with verbal conjunct and the adjunction errors into IP are special for Chinese, due to the subject *pro-drop*. Guo (2009) presented the other related work; Guo discussed the language-specific properties of Chinese, including the shortage of syntactic constraints, the *pronoun-dropping* and the *topic-prominence*.

In our work, we focused on the difficulties faced in Chinese deep parsing, and drew similar conclusions to the previous two related works. We revealed that the following three aspects brought difficulties to Chinese deep parsing: (1) the large distribution of Chinese verbs and their shortage of syntactic constraints; (2) the large variety of supertags for Chinese verbs, for which the subject *pro-drop* was considered to be the main reason; (3) the large numbers of relative

clauses and verbal coordination in Chinese, and the ambiguity in their analysis.

In addition to analyzing the parsing difficulty in Chinese deep parsing, some researchers focused on developing Chinese deep parsers.

Guo et al. (2007) built an LFG-based parser using wide-coverage LFG approximations induced from the Penn Chinese Treebank. This is the only previous work that had been conducted on Chinese deep parsing based on lexicalized grammars, although many related works had been done on English. Instead of training a parser based on the obtained LFG resources, Guo used an external PCFG parser to create c-structure trees, and then mapped the c-structure trees into f-structures using their annotation rules (Guo, 2009).

Besides of Guo's work, some researchers worked on joint dependency parsing and semantic role labeling to fulfill Chinese deep parsing (Li et al., 2010; Morante et al., 2009; Gesmundo et al., 2009; Dai et al., 2009; Lluís et al., 2009); other researchers focused on performing semantic role labeling after syntactic parsing (Fung et al., 2007; Sun and Jurafsky, 2004; Bjorkelund et al., 2009; Meza-Ruiz and Riedel, 2009; Zhao et al., 2009).

There were also some previous works that focused on building the language resources with lexicalized grammars, but not parsing with these resources. With the hand-crafted conversion rules, Yu et al. (2010) built a Chinese HPSG Treebank semi-automatically from the Penn Chinese Treebank. Guo (2009) also used rules to convert the Penn Chinese Treebank into LFG resources. Moreover, Tse and Curran (2010) built a Chinese CCGbank, which was also automatically induced from the Penn Chinese Treebank.

6 Conclusion and Future Work

In this paper, we discussed the prevalent difficulties in Chinese deep parsing, based on a lexicalized grammar theory – HPSG. All of the discussions were based on the analysis of a Chinese HPSG parser, which was trained on a Chinese HPSG Treebank, developed from the Penn Chinese Treebank. The analysis shows that since in Chinese, verbs have less syntactic constraints; the subject *pro-drop* appears frequently; furthermore, there is a larger distribution of ambiguous constructions, such as the relative clause and verbal coordination, deep parsing on Chinese is more difficult than on English. In addition,

compared with Chinese syntactic parsing, Chinese semantic parsing is more difficult, because of the inherent ambiguities caused by both verbal coordination and relative clauses.

To our current knowledge, it is the first work that makes a detailed analysis of the difficulty in Chinese deep parsing based on lexicalized grammars. The conclusions drawn in this work will be useful to other related works on Chinese deep parsing, by providing the possible future research directions. Moreover, the conclusions will also help us to improve the performance of the Chinese HPSG parser, by enhancing coordination disambiguation with the method proposed in (Kurohashi and Nagao, 1994); reducing the granularity of verb supertags, and so on. In addition, the Chinese HPSG parser, which had been applied in this work for comparison, will also be released this year.

References

- Anders Bjorkelund, Love Hafdell and Pierre Nugues. 2009. Multilingual Semantic Role Labeling. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared task*.
- Stephen Clark and James R. Curran. 2004. Parsing the WSJ Using CCG and Log-linear Models. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*.
- Qifeng Dai, Enhong Chen and Liu Shi. 2009. An Iterative Approach for Joint Dependency Parsing and Semantic Role Labeling. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL 2009)*.
- Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell and Annie Zaenen. 1995. *Formal Issues in Lexical-Functional Grammar*. Cambridge University Press, Stanford, CA.
- Ruth O'Donovan, Michael Burke, Aoife Cahill, Josef Van Genabith and Andy Way. 2005. Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks. *Computational Linguistics*: 329 – 366.
- Pascale Fung, Zhaojun Wu, Yongsheng Yang and Dekai Wu. 2007. Learning Bilingual Semantic Frames: Shallow Semantic Parsing vs. Semantic Role Projection. *Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007)*. 75-84.
- Yuqing Guo. 2009. *Treebank-based acquisition of Chinese LFG Resources for Parsing and Generation*. Ph.D. Thesis. Dublin City University.
- Yuqing Guo, Josef van Genabith and Haifeng Wang. 2007. Acquisition of Wide-Coverage, Robust, Probabilistic Lexical-Functional Grammar Resources for Chinese. *Proceedings of the 12th International Lexical Functional Grammar Conference (LFG 2007)*. 214-232.
- Jan Hajic, Massimiliano Ciaramita, Richard Johanson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared task*. 1-18.
- C. T. James Huang. 1989. Pro-drop in Chinese: A Generalized Control Theory. *O. Jaeggli and K. Safir (eds.). The Null Subject Parameter*. 185-214.
- Ronald M. Kaplan, Stefan Riezler, Tracy H. King, John T. Maxwell III and Alexander Vasserman. 2004. Speed and Accuracy in Shallow and Deep Stochastic Parsing. *Proceedings of Human Language Technology conference / North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL 2004)*.
- Sadao Kurohashi and Makoto Nagao. 1994. A Syntactic Analysis Method of Long Japanese Sentences based on the Detection of Conjunctive Structures. *Computational Linguistics*. 20(4): 507-534.
- Roger Levy and Christopher Manning. 2003. Is it Harder to Parse Chinese, or the Chinese Treebank? *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*.
- Junhui Li, Guodong Zhou and Hwee Tou Ng. 2010. Joint Syntactic and Semantic Parsing of Chinese. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*. 1108-1117.
- Xavier Lluís, Stefan Bott and Lluís Marquez. 2009. A Second-order Joint Eisner Model for Syntactic and Semantic Dependency Parsing. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL 2009)*.
- Mitchell P. Marcus, Beatrice Santorini and Mary Ann Marcinkiewicz. 1994. Building a Large Annotated Corpus of English: the Penn Treebank. *Computational Linguistics*. 19(2): 313-330.
- Takuya Matsuzaki, Yusuke Miyao and Junichi Tsujii. 2007. Efficient HPSG Parsing with Supertagging and CFG-filtering. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*.
- Lluís Marquez, Xavier Carreras, Kenneth C. Litkowski and Suzanne Stevenson. 2009. Semantic

- Role Labeling: An Introduction to the Special Issue. *Computational Linguistics*. 34(2): 145-159.
- Ryan McDonald, Kevin Lerman and Fernando Pereira. 2006. Multilingual Dependency Analysis with a Two-stage Discriminative Parser. *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*.
- Ivan Meza-Ruiz and Sebastian Riedel. 2009. Multilingual Semantic Role Labeling with Markov Logic. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared task*.
- Yusuke Miyao. 2006. *From Linguistic Theory to Syntactic Analysis: Corpus-oriented Grammar Development and Feature Forest Model*. Ph.D. Thesis. The University of Tokyo.
- Yusuke Miyao and Junichi Tsujii. 2008. Feature Forest Models for Probabilistic HPSG Parsing. *Computational Linguistics*. 34(1): 35-80.
- Roser Morante, Vincent Van Asch and Antal van den Bosch. 2009. A Simple Generative Pipeline Approach to Dependency Parsing and Semantic Role Labeling. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL 2009)*.
- Andrea Gesmundo, James Henderson, Paola Merlo and Ivan Titov. 2009. A Latent Variable Model of Synchronous Syntactic-semantic Parsing for Multiple Languages. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL 2009)*.
- Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gulsen Eryigit, Sandra Kubler, Svetoslav Marinov and Erwin Marsi. 2007 (a). MaltParser: A Language-independent System for Data-driven Dependency Parsing. *Natural Language Engineering*. 13(2): 95-135.
- Joakim Nivre, Johan Hall, Sandra Kubler, Ryan T. McDonald, Jens Nilsson, Sebastian Riedel and Deniz Yuret. 2007 (b). The CoNLL 2007 Shared Task on Dependency Parsing. *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*. 915-932.
- Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press and CSLI Publications, Chicago, IL and Stanford, CA.
- Mark Steedman. 2000. *The Syntactic Process*. The MIT Press.
- Honglin Sun and Daniel Jurafsky. 2004. Shallow Semantic Parsing of Chinese. *Proceedings of Human Language Technology conference / North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL 2004)*.
- Daniel Tse and James R. Curran. 2010. Chinese CCGbank: Extracting CCG Derivations from the Penn Chinese Treebank. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*.
- Nianwen Xue and Martha Palmer. 2009. Adding Semantic Rules to the Chinese Treebank. *Natural Language Engineering*. 15(1): 143-172.
- Nianwen Xue, Fei Xia, Fudong Chiou and Martha Palmer. 2005. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*. 11(2): 207-238.
- Kun Yu, Yusuke Miyao, Xiangli Wang, Takuya Matsuzaki, Junichi Tsujii. 2010. Semi-automatically Developing Chinese HPSG Grammar from the Penn Chinese Treebank for Deep Parsing. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*.
- Hai Zhao, Wenliang Chen, Chunyu Kity and Guodong Zhou. 2009. Multilingual Dependency Learning: a Huge Feature Engineering Method to Semantic Dependency Parsing. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared task*.