Semantic Labeling of Chinese Verb-complement Structure Based on Feature Structure

Bo Chen, Chen Lv, Xiaomei Wei, Donghong Ji

Computer School, Wuhan University, Wuhan, China

{chenbo,lvchen1989,xiaomeiwei,dhji}@whu.edu.cn

Abstract. Semantic relations of Chinese verb-complement structure are complicated, which is difficult to analyze semantic relations in NLP. This paper proposes a novel model based on "the Feature Structure theory" and applies it in representing the semantic relations among the four components, which are subject, verb, object and complement. We annotated the fifteen types of semantic relations, and compared the feature structure with the traditional dependency grammar. The results show that feature structure is an undirected recursive graph, which can describe more Chinese semantic information, and achieve higher annotating efficiency and accuracy.

Keywords: Chinese verb-complement structure, feature structure, semantic labeling, dependency structure, graph

1. Introduction

Semantic parsing is one of the most challenging tasks in natural language processing, as well as one of the main bottlenecks of large-scale applications of language information technology today [1], [2], [3]. Semantic parsing of Chinese phrases is a basic task. Because of the flexibility of word order, it is more challenging of Chinese semantic parsing. In natural language processing, it is urgent to solve the difficulties of semantic annotation of Chinese sentences, especially Chinese special structures, such as "verb-complement structure", "complex noun phrase", and so on.

As a typical Chinese special sentence pattern, verb-complement structure includes four elements, subject, verb, complement, and object. The semantic relations among the four parts are complicated, which include the semantic relations between the subject and the complement, the semantic relations between the predicate and the complement, the semantic relations between the object and the complement, and triple semantic relations among the subject, the verb, and the object[4], [5].

It is great significance to find a more effective method to annotate and parse Chinese verb-complement structure. This article studies a novel model for Chinese semantic representation based on Feature Structure, and analyzes fifteen types of semantic relations of Chinese verb-complement structure. Comparing traditional de-

pendency structure, we achieve a better result of semantic parsing based on Feature Structure.

2. Chinese Verb-complement Structure

1.1 Characteristics of Chinese Verb-complement Structure

Chinese verb-complement structure is unique to Chinese, which represent the complementary and supplementary semantic relations. Particle "de" is the form tag of the structure, such as: "lei de hen" (*means very tired*). There are also many direct combinations of words in Chinese, such as: "hao ji le" (*means very great*).

Chinese verb-complement structure can be described as: "Subject + Verb + Complement + Object". This paper focuses on the semantic relations of Chinese verb-complement structure. Based on the researches of linguistics, we provide the formal description and semantic annotation of the semantic relations of Chinese verb-complement structure. Fig.1 shows the types of semantic relations of the structure.

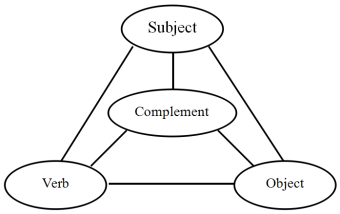


Fig. 1. The semantic relations of Chinese verb-complement structure

The semantic relations of Chinese verb-complement structure include two categories and six subtypes. One category includes the semantic relations between the complement and the subject, between the complement and the predicate, between the complement and the object. Another category includes the semantic relations among the subject, the predicate, and the object. Examples 1-3 are the typical sentences with verb-complement structure.

Example 1:

衣服 洗 干净 了。
/yifu/ /xi/ /ganjing/ /le/
Clothes wash clean
Clothes are washed clean.

Example 2:

衣服 洗 完 了。
/yifu/ /xi/ /wan/ /le/
Clothes wash finished
Clothes are washed up.

Example 3:

衣服 洗 晚 了。
/yifu/ /xi/ /wan/ /le/

Clothes wash late

Clothes are washed too late.

Syntactic structures of Example 1-3 are described as: “S (*clothes*) + V (*to wash*) + C (*adjective*)”. In Example 1-3, there are same subjects and predicates, only different complements which are adjectives. By contrast, we can find that even though the three sentences have same syntactic structures, the semantic orientations of the complements are different. In Example 1, the semantic orientation of the complement “ganjing” (*means clean*) is the subject “yifu” (*means clothes*). In Example 2, the semantic orientation of the complement “wan” (*means finished*) is the quantity of the subject “yifu” (*means clothes*). In Example 3, the semantic orientation of the complement “wan”(*means late*) is the time of occurrence of the predicate “xi” (*means to wash*). The similar syntactic structures of Chinese verb-complement structure bring difficulties of the accurate description of their internal semantic relationships.

1.2 Difficulties in Parsing Chinese Verb-complement Structure

Currently, traditional dependency structure is the main semantic analysis method to parse Chinese [6], [7]. Fig.2 is the dependency tree of Chinese verb-complement structure:

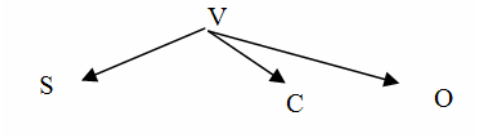


Fig. 2. the dependency tree of Chinese verb-complement structure

If we parse Example1-3 with dependency grammar, we will have the same dependency trees, as Fig.3.

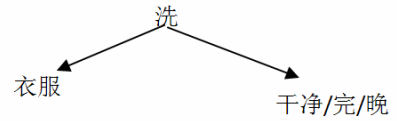


Fig. 3. the dependency trees of Example 1-3

Dependency grammar can not represent the semantic relations between the subject and the complement.

Example 4:

小王　　羞　红　了　脸。

/xiaowang/ /xiu/ /hong/ /le/ /lian/

Xiaowang shy red face

Xiaowang feels shy and his face turns t red.

Fig.4 is the dependency tree of Example 4.

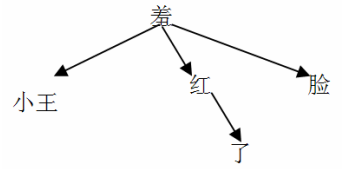


Fig. 4. the dependency tree of Example 4

According to semantic relatedness and semantic cognition, we analyze Example 4. There are 5 word pairs with semantic relation at least, such as:

[Xiaowang，xiu]；[Xiaowang，lian]；[xiu，hong]；[lian，hong]；[hong，le].

Fig.4 is the result of Example 4 using dependency grammar. There are only represented three semantic relations and omitted two semantic relations, which are the relations between the subject "Xiaowang" and the object "lian" (*means face*), and the relations between the object "lian" (*means face*) and the complement "hong"(*means red*). Using traditional dependency structure to parse Chinese verb-complement sentence will lost much semantic information, and will bring difficulties for the following Chinese processing.

2 Feature Structure Theory

Considering the unique characteristics of Chinese, we revised the traditional grammar and proposed the Feature Structure theory, which use feature triple to describe the semantic relation of every word pair[8], [9].

The final purpose of semantic parsing in Machine Translation is to find the semantic relations in a sentence [10]. We focus on the representation of semantic relations.

Example 5:

从	广州	飞	飞	到	武汉
/cong/	/Guangzhou/	/fei/	/fei/	/dao/	/Wuhan/
From	Guangzhou	to fly	to fly	to	Wuhan
Fly from Guangzhou			*Fly to Wuhan*		

Example 5 can be described as:

fei-cong (the beginning)- Guangzhou,
fei-dao (the destination)- Wuhan.

The two triples can be expressed as a set of triples of an Entity, a Feature and a Value:

Feature Triple: [Entity, Feature, Value]

It can represent a group of semantic relations. Example 5 can be described as following:

[fei, cong, Guangzhou]; [fei, dao, Wuhan].

In general, feature structure can describe the semantic relations and the type of the relations. A feature triple consists of entity, feature and its value. The triples can represent the semantic relations among multiple nodes. The formal representation of the triples is recursive and undirected graph. Fig.5 is the Feature Structure Graph of Example 5.

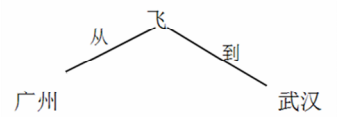


Fig. 5. the Feature Structure Graph of Example 5

Example 4 can be represented by Fig.6.

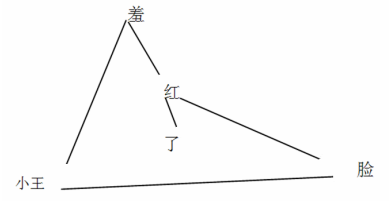


Fig. 6. the Feature Structure Graph of Example 4

In Fig.6, the multiple nodes are the subject "xaiowang", the complement "hong" (*means red*), the object "lian" (*means face*), and every node has semantic relations with at least two nodes. Feature structure can describe the entire semantic relations of Example 4.

Fig. 7 shows the Feature Structure Graph. Formally, Feature Structure can be described as a recursive undirected graph, which means that a node itself can be a graph, [8]. Feature Structure allows nesting and multiple correlations.

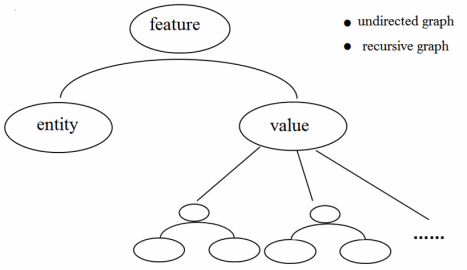


Fig. 7. the Feature Structure Graph

3 Comparing Feature Structure with Dependency Structure

Example 6:

他	喝	醉	了	酒。
/ta/	/he/	/zui/	/le/	/jiu/
He	to drink	drunk		alcohol

He is drunk.

The characteristics of the verb-complement structure of Example 6 are the complement "zui" (*means drunk*) has the semantic relation with the subject "ta" (*means he*). The sentence "ta he zui le jiu" can be simplified as: "ta zui le", which means that

the relation between the subject and the complement is subject-predicate. In Chinese, this category is popular. Example 6 includes five triples; Fig.8 is the Feature Structure Graph of Example 6.

[he, ,ta]；[he, ,jiu]；[he, ,zui]；[zui, ,ta]；[zui, ,le]

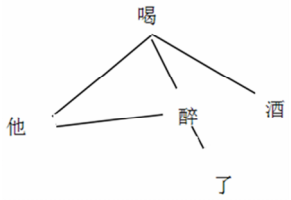


Fig. 8. the Feature Structure Graph of Example 6

Example 7:

一个月 踢 坏 了 三 双 鞋。
/yigeyue/ /ti/ /huai/ /le/ /sanshuang/ /xie/
One month tick broken three pairs shoes
He played football so hard to broke three shoes.

The characteristics of the verb-complement structure of Example 7 are the complement “huai” (*means broken*) has the semantic relation with the object “xie” (*means shoes*). The relation between the object and the complement is subject-predicate. In Chinese, this category is also very common. Example 7 includes five triples; Fig.9 is the Feature Structure Graph of Example 7.

[ti, ,yigeyue]；[ti, ,huai]；[huai, ,xie]；[xie, shuan, san]；[huai, ,le]

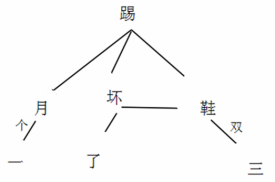
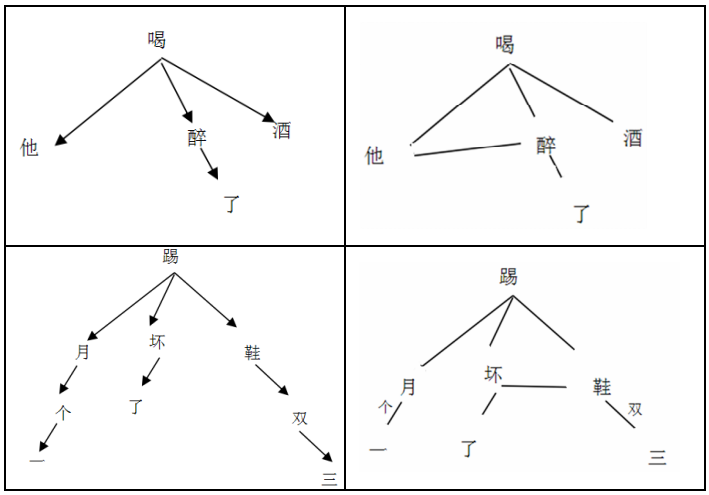


Fig. 9. the Feature Structure Graph of Example 7

We use two methods of Dependency Structure and Feature Structure to parse Examples 6-7. Table 1 is the Dependency Tree and Feature Structure graph of Example s 6-7.

Table 1. Dependency Tree and Feature Graph of Examples 6-7

Dependency Tree	Feature Structure Graph



In Table 1, Dependency Tree can not represent the semantic relations between the subject "ta" (*means he*) and the complement "zui" (*means drunk*), between the object "xie" (*means shoes*) and the complement "huai" (*means broken*). Dependency Tree lost two semantic relations.

As for Chinese verb-complement Sentences, Feature Structure graph can resolve the Chinese flexible word orders, and can describe the semantic relations among one word with more words, so Feature Structure graph can represent more semantic relations and more semantic information.

4 Conclusion and Future Work

We put forward a mechanism "Feature Structure", and annotated Chinese verb-complement structure with traditional dependency structure and Feature Structure. The results showed that Feature Structure could represent the complicated semantic relations of the four elements, and more other semantic information. It is used to represent Chinese phrases and sentences. Feature Structure can be represented as a recursive undirected graph, and allows nesting and multiple correlations.

It is an attempt to parse Chinese based on Feature Structure. So far we have built the basic concepts and description frameworks of Feature Structure, and built a large–scale Chinese semantic resource with 30,000 sentences. As for the applications, our research can be directly applied to relation extraction, event extraction, automatic question and answering as well as the syntactic parsing in machine translation.

5 Acknowledgment

The work is funded by the following projects: National Natural Science Foundation of China No.61202193, the Major Projects of Chinese National Social Science Foun-

dation No.11&ZD189, and the Chinese Postdoctoral Science Foundation No.2013M540593.

6 References

1. Chen, B. Ji, D., Chen, L.: Building a Chinese Semantic Resource Based on Feature Structure. International Journal of Computer Processing of Languages (2012) 95-101.
2. Chen, B. Wu, H., Chen, L.: Semantic labeling of Chinese serial verb sentences based on feature structure. Chinese Lexical Semantics (2013) 784-790.
3. Ji, D. Semantic annotation of Chinese phrases using recursive-graph. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, Hong Kong: Association for Computational Linguistics (2000)101-108.
4. Gao,Y. Semantic orientation of resultive construction in modern Chinese: a cognitive semantic perspective. Shanghai International Studies University (2005).[In Chinese]
5. Lv, S. Semantic analysis of sentences with resultive construction. the first international Chinese teaching conference(1985).[In Chinese]
6. Zhou, Q. Annotation Scheme for Chinese Treebank, Chinese Information Processing (2004) 1-8.[In Chinese]
7. Zhou, M, Huang CN. Approach to the Chinese Dependency Formalism For the Tagging of Corpus. Chinese Information Processing (1994)35-53.[In Chinese]
8. Chen, B. Ji, D., Chen, L.: Semantic Labeling of Chinese Subject-Predicate Predicate Sentence Based on Feature Structure. Journal of Chinese information processing(2012)24-32 [In Chinese]
9. Chen, B.: Building a Chinese Semantic Resource Based on Feature Structure. Doctoral Dissertation. Wuhan University, China, (2011) [In Chinese]
10. Feng, Z.: Machine Translation. Translation and Publishing Corporation (1998) 412-434[In Chinese]