Semantic Labeling of Chinese Serial Verb Sentences Based on Feature Structure

Bo Chen[1,3], Hongmiao Wu[2], Chen Lv[3], Hua Yang[3], and Donghong Ji[3]

[1] Dept of Language & Literature, Hubei University of Art & Science, Xiangyang, China
cb9928@gmail.com
[2] School of Foreign Languages & Literature, Wuhan University, Wuhan, China
Hongmiao23@163.com
[3] Computer School, Wuhan University, Wuhan, China
lvchen1989@gmail.com, yanghuastory@263.net, donghongji_2000@yahoo.com.cn

Abstract. Parsing Chinese serial verb sentences is a key issue in NLP. Many controversies arise from serial verb sentences. This paper puts forward a novel model "the Feature Structure theory" to resolve the semantic labeling of Chinese serial verb sentences. We analyze the difficulties in annotating these sentences, and compare Feature Structure with traditional dependency structure. Feature Structure represents more semantic information and more semantic relations. Feature Graph is a recursive undirected graph, allows nesting and multiple correlations.

Keywords: Chinese serial verb sentences, Feature Structure, semantic labeling, dependency structure, graph.

1 Introduction

In natural language processing, semantic parsing is one of the most challenging topics in the modern fields of computational linguistics, as well as one of the main bottlenecks of large-scale applications of language information technology today[1-3]. As a typical Chinese special sentence pattern, Chinese serial verb sentence has two and more verbs, which arises many problems in semantic parsing. This article studies a novel model for Chinese semantic representation based on Feature Structure, and analyzes Chinese serial verb sentence using Feature Structure. Comparing traditional dependency structure, we achieve a better result of semantic parsing based on Feature Structure.

1.1 Characteristics of Chinese Serial Verb Sentence

Chinese serial verb sentences can be described as: "Subject +Verb Phrase$_1$ +Verb Phrase$_2$ +Verb Phrase$_n$". Its characteristics include that two or more serial Verb Phrases compose the predicate of the sentence, they share a common subject. The structures of these verb

D.H. Ji and G.Z. Xiao (Eds.): CLSW 2012, LNAI 7717, pp. 784–790, 2013.

phrases are compact. There aren't syntactic relations between them, such as: subject-predicate, modifier-head, predicate-object, predicate-complement, etc. In terms of time sequence or logic sequence, these verb phrases are arranged in order [4-6].

Example 1:

我	开	车	去	车站	接	他
/wo/	/kai/	/che /	/qu /	/chezhan /	/jie /	/ta /.
I	drive	car	go	station	pick up	him

I drive a car to go to station to pick up him.

In Example 1, there are three verb phrases: “开车” (means drive a car), “去车站” (means go to station), “接他” (means pick up him). They are arranged by time sequence.

The semantic relations of a Chinese serial verb sentence include two parts: the semantic relations between the subject and these verb phrases, the semantic relations among these verb phrases.

1.2 Difficulties in Parsing Chinese Serial Verb Sentences

Currently, traditional dependency structure is the main semantic analysis method to parse Chinese language [7-8]. When we build Chinese labeling corpuses, most of us will choose dependency structure. However, when we use it to parse Chinese serial verb sentence, there are many difficulties.

In accordance with traditional dependency rule, only one verb can be the head of a sentence, all of the other words depend on it [1], [8]. As for Chinese serial verb sentence, there are at least two verbs, and maybe more. It is hard to ensure which one is the head.

Chinese information processing provides a compromise that the first verb of Chinese serial verb sentence is regarded as the head; the following verbs depended on it [9-10]. Fig.1 is the dependency tree of Example 1.

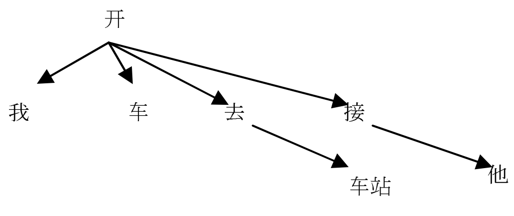


Fig. 1. Dependency tree of Example 1

According to semantic relatedness and semantic cognition, we analyze Example 1. There are 6 word pairs with semantic relation at least, such as:

(我, 开), (我, 去), (我, 接), (开, 车), (去, 车站), (接, 他).

Besides, Example 1 also contains other semantic information, we can ask questions to find it:

“where do we pick him up?”
“How do we pick him up?”
“How can we go there?”

The answers are “车站” (means station), “用车接” (means by car), “开车” (means drive).

Using traditional dependency structure to parse Chinese serial verb sentence will lost much semantic information, and will bring problems for the following Chinese processing.

2 Feature Structure Theory

The final purpose of semantic parsing in Machine Translation is to find the semantic relations in a sentence [11]. We focus on the representation of semantic relatedness and relatedness classification.

Example 2:

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

In Example 2, there are semantic relatedness between “飞” (means fly) and “广州 ”(means Guangzhou), between “飞”(means fly) and “武汉”(means Wuhan). If we apply relatedness classifications to Example 2, it can be described as:

飞-从(the beginning)-广州
飞-到(the destination)-武汉

The two triples can be expressed as a set of triples of an Entity, a Feature and a Value. We name it “Feature Triple” of the phrase or sentence structure [12].

Feature Triple: [Entity, Feature, Value]

A Feature Triple can represent a group of semantic relatedness. Example 2 can be described as following:

[飞, 从, 广州]
[飞, 到, 武汉]

Fig. 2 shows the graph of Feature Structure. Formally, a triple can be represented as two nodes and the edge connecting them. The nodes stand for words, the edge stands for feature. The node serves as the owner of the feature, while the other nodes as value.

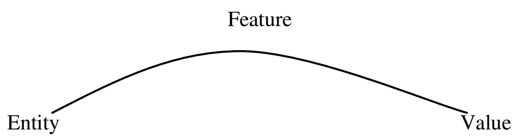


Fig. 2. Feature Structure

Example 3:

他	说	他	是	大学	教师
/ta/	/shuo/	/ta/	/shi/	/daxue/	/jiaoshi/
He	say	he	is	university	teacher

He says he is an university professor.

Example 3 can be described as following:

[说, ,他]

[说, ,他是大学教师]

[是, ,教师]

[是, , 他]

[教师, ,大学]

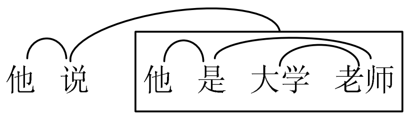


Fig. 3. Feature graph of Example 3

Fig. 3 shows that Example 3 has 5 feature triples. “他 (means he)” is the entity and one value of “说 (means say)”. “他是大学教师 (means he is a university professor)” is the other value of “说 (means say)”. Here “他是大学教师 (means he is a university professor)” is treated as an entirety to produce semantic relations with “说 (means say)”. And the node “他是大学教师 (means he is a university professor)” itself is a Feature Structure. “是 (means is)” is the entity, “大学教师 (means a university professor)” is its value, “他(means he)” is its another value. Besides, the node “大学教师 (means a university professor)” itself is also a Feature Structure. “教师 (means professor)” is the entity; “大学 (means university)” is its value.

Fig. 4 shows the Feature Graph. Formally, Feature Structure can be seen as a recursive undirected graph, which means that the node itself can be a graph [1], [8]. Feature Structure allows nesting and multiple correlations.

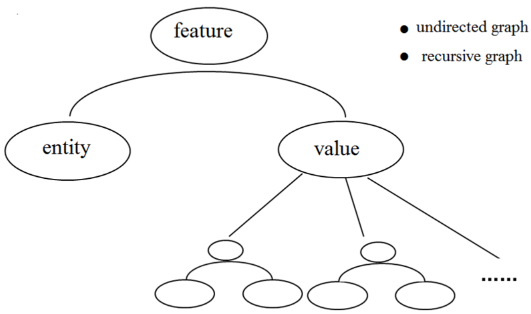


Fig. 4. Feature Structure Graph

3 Parsing Chinese Serial Verb Sentences with Feature Structure

Example 4:

我 推开 门 走 出去.
/wo/ /tui kai/ /men/ /zou/ /chuqu/
I push door walk outside
I open the door and go outside.

Example 4 is a typical Chinese Serial Verb Sentence. The first verb phrase is "推开门" (means open the door), the second verb phrase is "走出去" (means go outside), the two phrases are arranged in time order. Its feature graph is Fig.5:

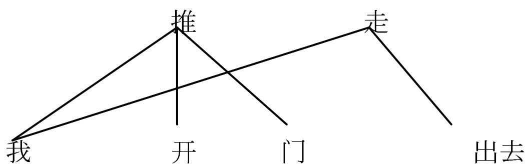


Fig. 5. Feature Graph of Example 4

Example 5：

我 买 了 碗 面 吃.
/wo/ /mai/ /le/ /wan/ /mian/ /chi/
I buy bowl noodle eat
I bought a bowl of noodle to eat.

In Example 5, the first verb phrase is "买了碗面" (means bought a bowl of noodle), "面" (means noodle) is the object of the verb "买" (means buy). The second verb phrase doesn't have object, only have a verb "吃" (means eat). The characteristics of Example 5 are the object of the first verb "买" (means buy)—— "面" (means noodle) is the patient of the second verb "吃" (means eat). The two verb phrases have semantic relatedness. Its feature graph is Fig.6[1]:

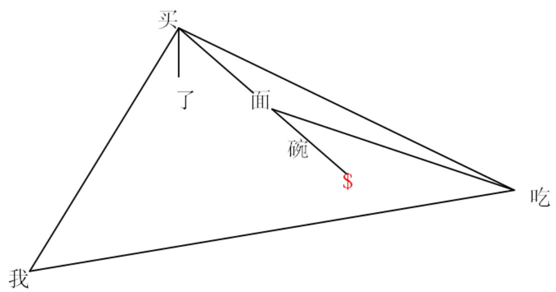


Fig. 6. Feature Graph of Example 5

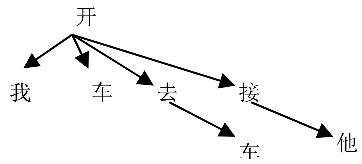
[1] The characteristics of Example 5 is that the number "一"(means one) in the phrase "一碗面" is omitted. we use"$"to describe the phenomena. It is the characteristics of Chinese language.

4 Comparing Feature Structure with Dependency Structure

We use two methods of Dependency Structure and Feature Structure to parse Example 5. Table 1 is the Dependency Tree and Feature Structure graph of Example 5, and Table 2 is the specific parsing results.

Table 1. Dependency Tree and Feature Graph of Example 5

Dependency Tree	Feature Structure graph



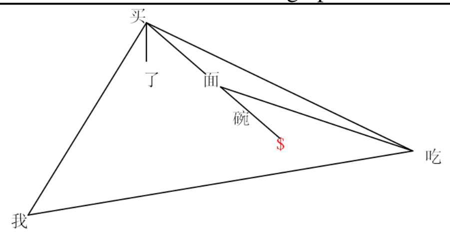


Table 2.

Semantic relations between words		The consequence of Feature Structure	The consequence of Dependency Structure
exist	Not exist		
我,买		+	+
买,了		+	+
买,面		+	+
碗,面		+	+
吃,面		+	-
我,吃		+	-
	买,吃	-	+

In Table 2, Dependency Tree can not represent the semantic relations between “我” (means I) and “吃” (means eat), “吃” (means eat) and “面” (means noodle). However, it represents the semantic relation between “买” (means buy) and “吃” (means eat) which does not exist. Dependency Tree omitted two semantic relations, and labeled a semantic relation that does not exist.

As for Chinese Serial Verb Sentences, Feature Structure graph can represent more semantic relations and more semantic information.

Firstly, Dependency structure can only describe the semantic relation between the subject and the first predicate verb. Feature Structure can describe completely the semantic relations between the subject and all predicate verbs.

Secondly, Dependency structure can not describe the semantic relation between the object of one predicate verb and another predicate verb. Feature Structure can describe it.

Thirdly, in Chinese Serial Verb Sentences, two predicate verbs may have semantic relations or not. Dependency Tree cannot discriminate the situation. No matter whether the actual semantic relation does exist or not, Dependency Tree will order the first verb as the head of the sentence. Feature Structure can represent the semantic relations in terms of the actual situation.

5 Conclusion and Future Work

According to the semantic representation, we put forward a mechanism "Feature Structure". 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 use Feature Structure to label Chinese Language. Now 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 used directly to relation extraction, event extraction, automatic question & answering as well as the syntactic parsing in machine translation.

Acknowledgment. The work is funded by the following projects: the Major Projects of Chinese National Social Science Foundation No.11&ZD189, National Natural Science Foundation of China No.61202193,61133012,61173062,61070082,61070083, and 61070243.

References

1. Chen, B., Ji, D.H.: Chinese Semantic Parsing Based on Dependency Graph and Feature. In: 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Shenyang, China, pp. 1731–1734 (2011)
2. Nivre, J., Scholz, M.: Deterministic dependency parsing of English text. In: Proceedings of the 20th International Conference on Computational Linguistics, pp. 64–70 (2004)
3. Nivre, J., Hall, J., Kübler, S., McDonald, R., Nilsson, J., Riedel, S., Yuret, D.: The CoNLL 2007 Shared Task on Dependency Parsing. In: Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL, pp. 915–932 (2007)
4. Li, L.D.: Modern Chinese Sentence Patterns, pp. 302–307. The Commercial Press (1986)
5. Xing, X.: on the characristics of Chinese Serial Verb Sentences. Journal of Xinjiang University, 116–122 (1987)
6. Yang, Y.R.: The semantic relation of Chinese Serial Verb Sentences and Chinese Pivotal Sentences. Journal of Southwest China Normal University, 96–100 (1992)
7. Buchholz, S., Marsi, E.: CoNLL-X Shared Task on Multilingual Dependency Parsing. In: Proc. of the 10th CoNLL-X, NY, pp. 149–164 (2006)
8. Chen, B., Ji, D.H., Lv, C.: Semantic Labeling of Chinese Subject-Predicate Predicate Sentence Based on Feature Structure. Journal of Chinese Information Processing, 24–32 (2012)
9. Zhou, M., Huang, C.N.: Approach to the Chinese Dependency Formalism For the Tagging of Corpus. Journal of Chinese Information Processing, 35–53 (1994)
10. Zhou, Q.: Annotation Scheme for Chinese Treebank. Journal of Chinese Information Processing, 1–8 (2004)
11. Feng, Z.W.: Machine Translation. Translation and Publishing Corporation, 412–434 (1998)
12. Chen, B.: Building a Chinese Semantic Resource Based on Feature Structure. Doctoral Dissertation. Wuhan University, China (2011)