

Semantic Labeling of Chinese Subject-Predicate Predicate Sentence Based on Feature Structure

Bo Chen^{1,2, a}, Chen lv^{2,b}, Donghong Ji^{2, c}

¹Dept of Language & Literature, Hubei University of Art & Science, Xiangyang, China

²School of Computer, Wuhan University, Wuhan, China

^acb9928@gmail.com, ^blvchen1989@gmail.com, ^cdonghongji_2000@yahoo.com.cn

Keywords: Feature Structure, semantic labeling, dependency structure, graph

Abstract. Labeling corpus with Feature Structure is a new attempt of the large-scale semantic descriptions and semantic annotation. This paper put forward the Feature Structure theory to resolve the semantic label of Chinese phrases and sentences. We chose “subject-predicate predicate sentence” as the research target in the paper, and summarized seven categories of the labeled graph. Feature Structure model contains more semantic information. In addition, feature structure can be used to account for some linguistic phenomena. Feature Structure is applied to tackle Chinese special sentence patterns and achieve a satisfactory effect.

Introduction

In natural language processing, semantic parsing is one of the most challenging topics in the modern fields of linguistics and computational linguistics, as well as one of the main bottlenecks of large-scale applications of language information technology today [1]. Among many problems of semantic parsing, semantic parsing of phrase and sentence is one of the most basic tasks.

This article studies semantic representation, resource building and parsing strategy of Chinese Sentences based on Feature Structure. Feature Structure is not a new term in the modern fields of linguistics and computational linguistics. Phonetics described syllables using a similar mechanism of Feature Structure long time ago. Later, the formal syntactic theories such as GPSG and LFG used the complexity of Feature Set to describe the syntactic structure, which is also similar to Feature Structure. In both cases, the definition of a set of features is used to distinguish between syllables and syntax structure, which has had a significant impact on the Generative Phonetics field. So far, we still have not seen the attempt of the large-scale semantic descriptions and semantic parsing by using Feature Structure. [2,3,4,5]

Background

The general semantic parsing needs to resolve three main issues: representation, the resources and parsing. We need to complete these three tasks. First of all, what kind of form can be used to represent the mechanism? Secondly, how can we establish the resources of phrases and sentences based on the mechanism? In addition, given a phrase or sentence, how can we deduce the form?

Traditionally, as to semantic parsing, there are two methods. They are syntactic structure parsing and dependency structure parsing [1,6]. However, there are several problems if we use either of methods to analyze the Chinese language.

Firstly, current syntactic category is only suitable for typical syntactic structures. As for some special structures in Chinese, such as serial noun phrase and verb-complement phrase, it's difficult to distinguish the syntactic features among the words. In a serial noun phrase, each word is a general noun; it's too hard to ensure the different syntactic features.

Secondly, dependency structure only represents the dependency relations between two words. The non-head always depends on the head in a sentence. Dependency grammar is usually assumed mono-dependency, which means that a node can only depend on another node. Dependency grammar implies a hypothesis that the status of the words is inequitable. However, as to some special structures in Chinese, such as the serial noun phrase, there are some exceptions. Most often

than not the grammatical status of noun phrases are equal, so we may assume that no heads exist between serial noun phrases. Sometimes there are multiple dependency relations, namely, a node can be dependent on more than two nodes.

Hence, traditional syntactic structure parsing and dependency structure parsing are not suitable for Chinese semantic analysis. We put forward the concept of Feature Structure to resolve it.

Feature Structure theory

Generally, a phrase or sentence can be expressed as a set of triples of an entity, feature and value. We call the set “Feature Structure” of the phrase or sentence structure.

Example 1:

1)/hong/ /yanse/ /qiche/ 2)/hong/ /qiche/
red car red color car

In 1), “qiche(car)” is the entity, “yanse(color)” is a feature of “qiche(car)”, “hong(red)” is the value of the feature “yanse(color)”. “yanse(color)” links “qiche(car)” one side, and in the meantime, it links “hong(red)” to the other side. So “yanse(color)” can be used to be the tag of the connection of “qiche(car)” and “hong(red)”. In 2), “qiche(car)” is the entity, “hong(color)” is the value. We can express them as triples:

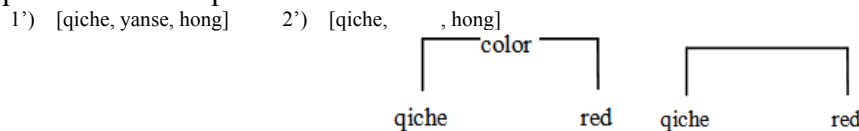


Fig.1 the Feature Structure Graphs of 1) & 2)

In 1), all of the entity, feature and value come out in the phrase. But in 2), the feature of “qiche(car)” is hidden. Only the entity and value come out.

Example 2:

/ta/ /shuo/ /ta/ /shi/ /daxue/ /jiaoshi/
He says he is university professor.

The triple is:

[shuo, ,ta], [shuo, ,tashidaxuejiaoshi], [shi, ,daxuejiaoshi], [jiaoshi, ,daxue], [shi, ,ta]

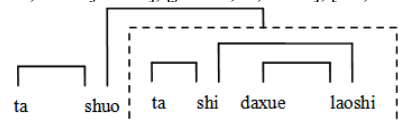


Fig.2 the Feature Structure Graph of Example 2

We can find from example 2: feature and value can be used as entities and represented in the Feature Structure. “ta(he)” is the agent and a value of “shuo(say)”. “ta shi daxue jiaoshi(he is a university professor)” is the other value of “shuo(say)”. Here “ta shi daxue jiaoshi(he is a university professor)” is treated as an entirety to make semantic relations with “shuo(say)”. And the value “ta shi daxue jiaoshi(he is a university professor)” itself is also a Feature Structure. “shi(is)” is the entity, “daxue jiaoshi(a university professor)” is its value, “ta(he)” is its another value. Furthermore, the node of the value “daxue jiaoshi(a university professor)” itself is also a Feature Structure. “jiaoshi(professor)” is the entity, “daxue(university)” is its value.

Formally, a triple can be represented as two nodes and the edge connecting them, in which the nodes stand for words, the edge stands for feature. The feature must be a feature of a node. The node serves as the owner of the feature, while the other nodes as value. Thus a Feature Structure can be represented as an undirected graph. Taking into account the value can also be another Feature Structure, so Feature Structure can be seen as a recursive graph, which means that the node itself can be a graph [7].

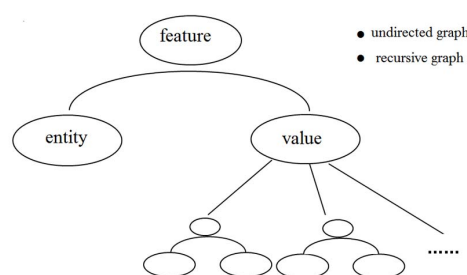


Fig.3 the Feature Structure Graph

In a word, compared to syntactic structure, Feature Structure is similar to dependency structure. Feature Structure mainly describes the relations among words, so that we do not need to define the syntactic category. Even in the recursive Feature Structure, we do not need to define the type of Feature Structure. Compared to dependency structure, on one hand, Feature Structure allows nesting, on the other hand, it also allows multiple correlations [2,3,4,5].

Our research

In the paper, we choose Chinese Subject-predicate Predicate Sentence as the research objective. The sentence with Subject-predicate phrase served as its main predicate is called Subject-predicate predicate sentence. It is one of the typical sentence patterns in Chinese language, which is the subordinate pattern of Chinese Subject-predicate sentence. Even though there are plenty of articles about Chinese Subject-predicate sentence by adopting linguistic ontological methods. The earliest studies can be traced back to 1921. However, 80 years up till now, no defined criterion has been available to analyze Subject-predicate predicate sentence in the field of linguistics. That is why the representation of Subject-predicate predicate sentence held the bottleneck of Chinese information processing.

In linguistic field, Subject-predicate predicate sentence can be described as: “N1 + N2 + V/A + N3”. N1 is the subject of the whole sentence, while “N2 + V/A + N3” is the predicate of the whole sentence. N1 is also called the “Big Subject”, which is the subject of the whole sentence. N2, which refers to the subject of Subject-predicate phrase that served as the predicate of the whole sentence, is also called “Small Subject”, which is the subject of subject-predicate sentence; V/A refers to the predicate of the Subject-predicate phrase, which can be served as adjectives, verbs, etc. N3 refers to the object of the Subject-predicate phrase.

Subject-predicate predicate sentence falls into three categories and thirteen subcategories in [8]. We labeled the semantic relations of the corpus of Subject-predicate predicate sentence on the basis of Feature Structures, and summarized seven categories of the labeled graphs. Take 3) for example:

3):

/women/ /shui/ /ye/ /libukai/ /shui/

We can't live without each other.

[libukai, ,shui1], [libukai, ,shui2], [libukai, ,women], [women, ,shui1], [women, ,shui2]

The category of Example 3) is:

[V, ,N1], [V, ,N2], [V, ,N3], [N1, ,N2], [N1, ,N3]

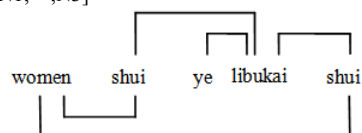


Fig.4 the Feature Structure Graph of Example 3

In Fig.4, “libukai (can't live without)” is the entity, “women(we)” “shui(who)” and “shui(whom)” are the three values of “libukai (can't live without)” respectively. In next recursive graph, “women(we)” is the entity, “shui(who)” and “shui(whom)” are its value. The relationship among “women(we)” “shui(who)” and “shui(whom)” is “optional reference” or “nonspecific reference”.

Labeling the Chinese sentences using Feature Structure, we may find it is very easy to determine which components are the entities; which components are the features and which components are the value. The Semantic relations among them are expressed clearly.

Conclusion and Future Work

Regarding the semantic representation, we put forward a new mechanism “Feature Structure”. It is used to represent Chinese phrases and sentences. “Feature Structure” allows recursive. Many semantic relations in Chinese sentences are not between words, but between word and a clause, or between a clause and a clause. In some special semantic relations is among three constituents. “Feature Structure” can resolve the problems.

The use of Feature Structures to label the semantic relations of Chinese sentences is an attempt to

build Chinese Semantic Resources. We achieved good results. Compared to dependency structure, Feature Structure can represent more semantic relations and allow multiple links. Regarding the semantic resources, we have built a large-scale corpus of Feature Structures with 30,000 raw sentences from Penn Chinese Treebank and Chinese press corpus. It enriches Chinese Semantics resources to some extent. Regarding 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: Resources Construction and Automatic Analysis on Chinese Feature Structure, Key Research Plan of NSF Grant No. 90820005; Semantic Description and Application on Chinese Feature Structure. Project 985, Grant No. 985yk004; Using Complex Network for Measuring Lexical Semantic Relatedness”, NSF Grant No.61070243.

References

- [1] Z.W.Feng: Machine Translation.China Translation and Publishing Corporation(1998), p.412.
- [2] B.Chen and Y.J. Ma: On labeling with dependency relations, Proc.ICCC 2007. Singapore: Chinese and Oriental Language Information Processing Society(2008), p. 205-208.
- [3] B.Chen, D.H.Ji: Chinese Semantic Parsing Based on Dependency Graph and Feature.Proc.EMEIT.China(2011), p.1731.
- [4] B.Chen, C.Lv: Semantic Labeling of Chinese serial verb sentences Based on Feature Structure. Proc. CLSW. China(2012). in press.
- [5] B.Chen, C.Lv: Building a Chinese Semantic Resource Based on Feature Structure. IJCPOL(2012). in press.
- [6] J. Nivre, J. Hall, S. Kübler, R. McDonald, J. Nilsson, S Riedel and D Yuret: The CoNLL 2007 Shared Task on Dependency Parsing.Proc. EMNLP-CoNLL.Prague(2007), p. 915.
- [7] D.H.Ji: Semantic Annotation of Chinese Phrases Using Recursive-graph. Proc. 38th ACL, Hong Kong(2000), p.101.