Classified Description and Application of Chinese Constitutive Role

Mengxiang Wang
Teachers' college
Beijing Union University
Beijing, China
e-mail:wmx1314-@126.com

Cuiyan Ma
Research Institute of Information Technology
Tsinghua University
Beijing, China
e-mail:mcymcy402@163.com

Abstract— **Constitutive role is one of the 4 qualia roles , which expresses a kind of constitutive relationship between nouns. According to the original definition and description characteristics, this paper divides the constitutive roles into two categories: materials and components. At the same time, combined with the previous methods of extracting the role automatically, this paper optimizes the method of extracting the role automatically. Relying on auxiliary grammatical constructions, we extract noun-noun pairs from large-scale corpus to extract descriptive features of constitutive roles, and then classifies these descriptive knowledge by manual double-blind proofreading. Finally, the author discusses the application of Chinese constitutive roles in word-formational analysis , syntactic analysis and synonym discrimination.**

Keywords- constitutive role; qualia role; classified description

I. INTRODUCTION

Constitutive role was proposed by the Pustejovsky in describing the qualia role structure of nouns.The qualia role structure is one of the core content of the Generative Lexicon (GL) Theory. (Pustejovsky 1991). The constitutive role is mainly to describe what is X made of / from, which is a relationship between a noun and another noun.

When the GL Theory was introduced into China, because of the difference between Chinese and English grammar system, some Chinese scholars have different understanding of the GT theory, including the number and definition of qualia structures. i.e In the beginning, when the Pustejovsky introduced the qialia structure, there were 4 qualia role: Constitutive, Agentive, Formal and Telic. Later according to the characteristics of Chinese grammar the Chinese scholar Yuan (2014) proposes a descriptive system of Chinese noun's qualia structure, which defines ten qualia roles (Constitutive, Orientation, Evaluation, Material, Formal, Agentive, Action, Telic, Handle). Yuan(2014) believed that besides the internal constitutive property, the constitutive roles can describe the size, shape, dimension, color and orientation of objects, and even include the arguments arguments involved in nouns. For example, the constitutive role of "opinion" includes "presenter, targeted matter",while the constitutive role of "mother" includes "address person, be addressed person". Yuan's description of constitutive roles contain a special purpose,but it is too complex, which is not in line with Pustejovsky's original idea of describing constitutive role. At the same time, it is difficult for extracting the relevant language knowledge, so it is not suitable for language processing.

In view of the characteristics of constitutive roles, this paper will classify the constitutive roles internally, and optimize the method of acquisition of constitutive roles by predecessors. Last but most important, we discuss their applications in word-formational analysis, syntax analysis, synonyms discrimination.

II. THE CLASSIFIED DESCRIPTION OF CONSTITUTIVE ROLE IN CHINESE NOUN

From the original idea， the constitutive role is to describe what is X made of. i.e, If the house is made of wood, the noun "wood" is the constitutive role of "house". But Chinese people's cognitive habits are different from the West, "made of" and "made from" is the same in Chinese. Therefore, we must classify the Chinese constitutive roles.

In fact, According to Pustejovsky's description of the concept of the constitutive role, the relationship of the constitutive roles is similar to the "part-of". Odell(1998) proposed that the "part-of" relationship contains six forms of expression.

As we can see from the table1, the types of "part-of" is still complex. From the point of view of applications, such a subdivision is not necessary. We see that apart from "Material-objects" type, the rest can be classified into one category: "Components-objects". Because the chair legs can be seen as a component of chair. The relation of "car wheel and car", "state and country", "display and computer", "husband and couple" all can be seen as "components-objects". Based on this, we merge the six forms of composition into two types: materials and components. The corresponding expressions of our constitutive roles are also semantic. We can describe the constitutive roles of a noun as The following two types:

(1) 桌子[木头, 石头, 竹…]
(desk[wood, stone, bamboo…]).
Material-objects type
(2) 电视机[电路板，显示屏，喇叭…]
(television [circuit board, screen, horn…]).
Components-objects type

III. THE EXTRACTION OF CONSTITUTE ROLE OF CHINESE NOUNS

As the extraction and description of Chinese qualia roles are in the initial stage, the extraction of Chinese qualia roles is basically based on artificial extraction. At present, only Wang(2016) has made an attempt to automatically extract the constitutive role.

Table 1: Odell's classification of "part-of"

Representation of "part-of"relationships	Annotation	Example
Component-integral object	An object consisting of a necessary component	The chair is composed of a chair legs, a chair surface, and a chair back. (Chair legs, chair surface, chair back are all necessary parts of the chair)
Material-object	An object consisting of a necessary material	The chair is made of wood.(wood is the necessary by making chair)
Portion-object	An object mostly consisting of portion	A car is made up of wheels and a frame. (The wheels and frame are most of the components of the car, but the car has other devices besides the wheels and frame.)
Place-area	An area is made up of places or location	China was made up of nine states including jizhou, qingzhou, jingzhou.(the jizhou, qingzhou, jingzhou consist of the whole china, in ancient times)
Member-bunch	An object consisting of a collection of members (member changes do not affect the whole unless the collection of members changes).	The computer consists of a host and a display.(the host and the display are a collection, including network cards, graphics cards, and other components
Member-partnership	An object consisting of a collection of members' partner.(the change of the member will affect the whole)	Couples are composed of husband and wife(husband and wife are interdependent and become couple)

On that time, Wang(2016) did not take into account the fact that the pattern characteristics of the constitutive role, only just use "由……构成（be consist of）" pattern, which can ensure the accuracy of extracting, but many information were missing. Therefore, this paper optimizes on the basis of its method. Based on more specific keywords and syntactic configuration matching, we extract specific noun pairs by the shortest path rule, and then extend the synonyms to obtain the knowledge of the constitutive roles. The specific steps are as follows:

Steps 1 pattern recognition

First, we use special words and syntactic constructions to match the desired pairs of nouns from massive texts. This process is called pattern recognition. The special words, syntactic constructions and patterns are shown in the table2.

Table 2: Key Words and Syntactic Constructions In the consititutive role pattern recognition

Special words or constructions	**Example**
有 (have/there be)	安徽 有 黄山 Anhui there is Huangshan Mountain "There is a Huangshan Mountain in Anhui "
构成/组成 (Constitute/consists)	这支军队 由 农民 组成 Army peasantsconsists "The army consists of peasants"
分为 (divide)	襄樊 分为 襄城 和 樊城 Xiangfan devide Xiang city and Fan city "The Xiangfan is divided into Xiang city and Fan city."
用……做(be madeof/ from)	这酒 是 用 高粱 做的 Wine is Sorghum made "This wine is made of sorghum."
比（例）如 such as	我 喜欢 吃 北方菜，比如 锅包肉、地三鲜 I like eat northern dishes, such as pot-fried meat, Disanxian. "I like to eat northern dishes, such as pot-fried meat, Disanxian."
偏正结构 modifier-headword NP+NP Combinations	荞麦饼 Buck wheat cake "a kind of cake that is made of buckwheat"

The single-line nouns are the consititutive roles of double-line nouns. It needs to be noted that some patterns are quite generic and using them to extract information can result in false positives (like the "have" pattern), so we use artificial methods to distinguish internal differences such as "have" and "there be" to avoid the noise.

Steps 2 collect (n_i-n_j) nouns pairs. To comb and clean up the extracted nouns pairs according to their co-occurrence rate.

This is mainly due to the fact that once the matching pattern increases, it will cause more data sparse. For example, the semantic complexity covered by "有（have）" is much greater than the semantic relationship of the constitutive characters. But "有（have）" as a keyword can match many compositions, so on the one hand, we have to spread it smoothly, and on the other hand we have to carefully extract. We think that co-incidence(CoR) is a relatively simple and effective method. A certain number of (n_i- n_j) co-occurrence may indicate a semantic link between two nouns, for example a qualia link. So we calculate the correlation between the nouns (n_i- n_j) from the Chinese GigaWord Corpus and SogouCA Corpus by following equation 1, and rank the unlabelled element (n_j) of (n_i) .

$$\mathrm{CoR}(n_i, n_j) = \frac{2*\mathrm{Sum}(n_i, n_j)}{\mathrm{Sum}(n_i)+\mathrm{Sum}(n_j)} \qquad (1)$$

Sum(n_i, n_j) presents co-occurrence number of (n_i, n_j). Sum(n_i) and Sum(n_j) means the number of n_i and n_j.

Steps 3 Extract the shortest path pattern of the matched (n_i-n_j)nouns from the corpus and to determine the (n_i-n_j)nouns set with constitutive relationship.

Steps 4 Through Baidu search engine, expanding the collection of constitutive roles of (n_i-n_j)nouns.

Steps 5 To extend the knowledge of the constitutive roles through the synonym set

Steps 6 Binary classification of (n_i-n_j) nouns with constitutive relationship.

Steps**3,4,5** can refer to Wang(2016) in detail. For steps **6,** we use artificial double-blind notation to classify all noun terms into 2 types(Components-objects or Material-objects), which contained constitutive role. In the process of labeling, in addition to judging directly from semantics, we also use some syntactic patterns to assist in judgment. For example, pattern "composed of ...（由……构成）" is generally classified as a Components-objects class.

IV. THE APPLICATIONS OF CONSTITUTIVE ROLE OF CHINESE NOUNS

4.1 Application in word-for mational analysis

Noun-noun compound phrases are sometimes not included in dictionaries. What is more, the grammatical type of these words are complex. For example: "石头桌子（stone desk）" means "the desk is made of stone", while "中国桌子（china desk）" means "Chinese desk". Song Peijie (2014)pointed out there are 3 types of the grammatical structure patterns of the noun-noun compounds: head word-modifier pattern, modifier-headword pattern and parataxis pattern. We find that if the N_1 and N_2 are constitutive relationship, the grammatical structure of "noun+noun compounds" are only 2: headword-modifier pattern and modifier-headword pattern. For the analysis of these "noun+noun compounds", Song Peijie (2014) pointed out that "implied predicates are the key to the analysis of the internal semantic relations of noun-noun compound words". Weixue (2013) point out that each of Noun-Noun Compounds contains a event, and the event is triggered by implicit predicates. We found that if theN_1 is N_2's constitutive role, the implicit predicates are almost "make", "constitute", "divide". We can make the following inductions:

If $N_1$is the constitutive role of N_2, the NN structure can be decomposed as follows:

(1)N_1+N_2 =N_1+v(做/组成（make or constitute）)+N_2

If N_2 is the constitutive role of N_1, the NN structure can be decomposed as follows:

(2)N_1+N_2=N_1+的(de)+N_2

For example, to "石头桌子（stone desk）", "石头（stone）" is the constitutive role of "桌子（desk）", so the structure can be decomposed as "石头做的桌子（stone+做（make）+桌子（desk））", which use the Form (1).

To "人脸（people face）", "脸（face）" is the constitutive role of "人（people）", so the structure can be decomposed as "人（people）+的（de）+脸（face）", which use the Form (2).

We have extracted 4610 disyllabic non-juxtaposed noun-noun compounds from the Modern Chinese Dictionary (2012 edition) and found that there are 2143 compounds which N_1andN_2 are constitutive relationship. Among them, 2012 compounds belong to "N_1 is N_2's constitutive role" type, of which 1822 phrases (90.6%) can be interpreted in the form of (1). But there are only 131 compounds belong to "N_2 is the constitutive role of N_1" type, of which 125 phrases (95.4%) can be interpreted in the form of (2).

4.2 Application in syntactic analysis

The structure "S+V1+O1+V2+O2" is usually analyzed in 2 ways:

(A) S+V1+(O1+V2+O2);

(B) (S+V1+O1)+V2+O2;

It is difficult to distinguish it through the existing NLP software.

But if the O2 is O1's constitutive role, the"(O1+V2+O2)" usually cannot exist (Unless the V2 is relation verb or certain verb). We can use this feature to analysis the sentences more correctly. The following figure can express this process:

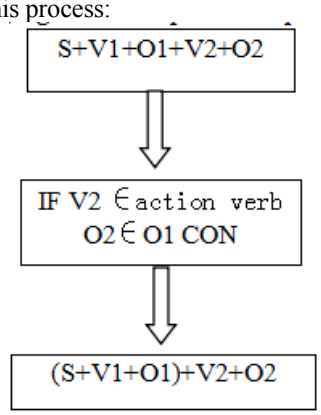

Figure 4: Automatic syntactic analysis of "S+V1+O1+V2+O2"

For example, We can see these 2 sentences:

(1)美国警告俄罗斯将插手委内瑞拉内政

USA warn Russia will intervene in Venezuela's internal affairs

*(1a)USA warning: Russia will intervene in Venezuela's internal affairs***[S+V1+(O1+V2+O2);]**

(1b)USA warns of Russia, the USA will intervene in Venezuela's internal affairs. **[(S+V1+O1)+V2+O2]**

(2)美国 警告 委内瑞拉 将 颠覆 马杜罗政权

*USA warns of Venezuela, the USA will subvert Maduro regime***[(S+V1+O1)+V2+O2]**

In the sentence (2), S is USA, V1 is warns, O1 is *Venezuela*, V2 is *subvert*,O2 is *Maduro regime*, V2 is action verb, and *Maduro regime* can be considered as the constitutive role of "*Venezuela*". We must output "(S+V1+O1)+V2+O2" automatically.

But if we don't tell the computer this constitutive knowledge, the result will be another one.

We use the Language Technology Platform(developed by Harbin Institute of technology, China) to analyze the syntactic structure of the Chinese sentence, the result is in Figure 5:

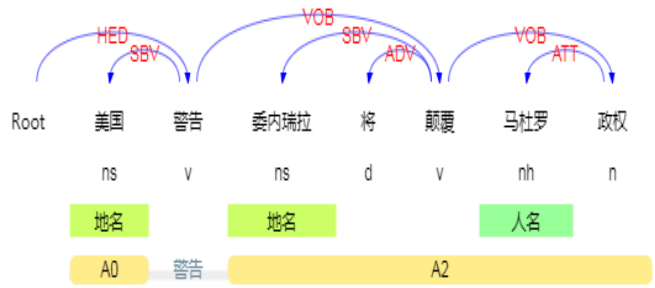

Figure 5: Syntactic analysis results by machine by Harbin Institute of technology, China

From the analysis chart of this platform, we can see that the actors of "颠覆（subvert）" are labelled "委内瑞拉（Venezuela）", not the "美国（USA）". If we add the constitutive knowledge of nouns, this mistake can be avoided. In this case, the constitutive role feature can help us to improve the syntactic analysis result.

4.3 Application in synonyms discrimination

Traditional synonyms discrimination is usually differentiated by sentence interpretation or sememe analysis, which usually focuses on the rational meaning of words. But apart from the rational meaning, it is possible that ordinary people pay more attention to the usage of words, which will involve grammar and pragmatics.

As one of descriptor of the noun ontological meanings, the constitutive role is contained in every noun. We find that if we can construct a complete system of constitutive role's description, We can distinguish some noun synonyms much better.

For example, in our constitutive role's description system, we describe the "奥秘(mystery)" vs "秘密(secret)" like this:

(1)The constitutive role of "奥秘 (mystery)":

The universe (mystery), the stars (mystery), the sun (mystery), natural phenomena (mystery), Luminescence (mystery), physiology (mystery), microorganism (mystery), heredity (mystery), etc.

(2)The constitutive role of "秘密(secret)":

Economy (Secret), Military (Secret), Diplomacy (Secret), Politics (Secret), Commerce (Secret), Nuclear (Secret), State (Secret), Communication (Secret) etc.

From the constitutive knowledge of the "奥秘(mystery)" and "秘密(secret)", we can find that "奥秘(mystery)" often describe the object things in nature, while the "秘密(secret)" tends to describe man or man-made results.

In this way, we can analyze the differences of synonyms very concretely, especially to the usage environment.

V. CONCLUSION

This paper just made a classified description of constitutive roles of Chinese nouns, improved an approach for the acquisition of constitutive role for Chinese nouns based on predecessors, and then discuss edits application in word-formational analysis, syntax analysis, synonyms discrimination. In the future, we will discuss how to expand the specific application of constitutive role in nature language processing.

ACKNOWLEDGMENT

This research was partly supported by National Science Foundation of China(No.61602040)and Supporting Plan for Cultivating High Level Teachers in Colleges and Universities in Beijing(CIT&TCD201904072)

REFERENCES

[1] Cimiano P and Wenderoth J. 2. Automatic Acquisition of Ranked Qualia Structures from the Web. In Pro-ceedings of the Meeting of the Association for Computational Linguistics, 2007, pages:888--895.

[2] Claveau V and Sébillot P. Automatic Acquisition of GL Resources, Using an Explanatory, Symbolic Technique. Advances in Generative Lexicon Theory, 2013, pages 431-454.

[3] E.M. Voorhees. Query expansion using lexical semantic relations. In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, pages 61–69.1994

[4] Gonzalez M, Lima V L S D. Redefining Traditional Lexical Semantic Relations with Qualia Information. Revista Palavra, 2004.

[5] Huang, Chu-Ren and Adam Kilgarriff, et al.. Chinese Sketch Engine and the Extraction of Collocations. In Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing. 2005

[6] LI Qiang. Semantic Analysis and Construe of Adjective-noun Composition Based on the Qualia Structure. Chinese Language Learning,2014

[7] Li Qiang and Yuan Yulin. A Qualia role Based Approach to the Analysis of Chinese Synonymous Nouns. Chinese Teaching in the world,2014, 04: pages 519-531.

[8] Monachini M, Moruzzi V, and Quochi V, et al. Lexical Relations and Domain Knowledge: The BioLexicon Meets the Qualia Structure.2007

[9] Pustejovsky. The generative lexicon. Computational Linguistics, 1991,17(4): pages 209–441.

[10] Pustejovsky, P. Anick, and S. Bergler. Lexical semantic techniques for corpus analysis. Computational Linguistics, Special Issue on Using Large Corpora II,1993, 19(2): pages 331–358.]

[11] Song Peijie Study on the Chinese noun-noun compounds Doctoral dissertation Of Huazhong university of science and technology ,2014

[12] Voorhees. Query expansion using lexical semantic relations. In Proceedings of the 17th annual international ACMSIGIR conference on Research and development in information retrieval, 1994,pages 61–69.

[13] Weixue and Yuan Yulin Constructing the Interpretation Templates for Chinese Noun Compounds Based on the Semantic Classes and the Qualia Roles of the Component Nouns, Chinese Teaching in the World,2013

[14] Wang Mengxiang,Wang Houfeng and Zheng Lvexing, The extraction and application of constitutive role of nouns in Chinese,IALP2016

[15] Yuan Y. On a descriptive system of qualia structure of Chinese nouns and its application in parsing complex Chinese grammatical phenomena. Contemporary Linguistics,2014

[16] Zeng, D., Liu, K., and Lai, S. et al. Relation classification via convolutional deep neural network. In Proceedings of COLING, pages: 2335–2344,2014

[17] Zheng Lvexing, Wang Houfeng, and Lv Xueqiang. Improving Chinese Dependency Parsing with Lexical Semantic Features. Natural Language Processing and Chinese Computing ,2015