Representation of Original Sense of Chinese Characters by FOPC

Yajun Pei

Communication University of China,100024
peiyajun2004@yahoo.com.cn

Zhiwei Feng

Ministry of Education Institute of Applied Linguistics,100010
zwfengde@bbn.cn

Abstract. In Natural Language Processing(NLP), the automatic analysis of meaning occupies a very important position. The representation of original sense of Chinese character plays an irreplaceable role in the processing of advanced units of Chinese language such as the processing of syntax and semantics, etc. This paper, by introducing a few important concepts: FOPC, Ontology and Case Grammar, discusses the representation of original sense of Chinese character.
The Key words : FOPC, Ontology, Case Grammar

1   Instruction

In Natural Language Processing(NLP), the automatic analysis of meaning occupies a very important position. This paper, by introducing a few important concepts: FOPC, Ontology and Case Grammar, discusses the representation of original sense of Chinese character.

2   The interpretative features of original sense of Chinese character dictionary

Before the establishment of the theory of Chinese characters' structure, Chinese character study belongs to the domain of explanation theory. Those who were engaged in this study were mostly researchers who study ancient Chinese. The theory of ancient Chinese characters is a type of an abstruse knowledge. Thus, to most of researchers, the threshold may be too high. Besides, the attributes of the Chinese characters stipulates that ancient Chinese is not the integrated unit of sense in Chinese. This can give us an appropriate interpretation about why the research achievement of characters can not be combined with the latest linguistic theories and become a deserted area by the information processing. The theory of the formation and structure of Chinese character interpreters Chinese characters using a brand new way, bringing the Chinese characters research into the modern track for the first time, and started the new era of the Chinese characters research.

The Chinese characters' sense is mainly divided into original sense and derived sense. we here only consider the original sense of Chinese characters. In addition, we think, although a Chinese character doesn't represent a smallest unit of sense, there is an unavoidable relationship between the sense of the Chinese characters and the sense of the words, phrase that constituted by it. And, from the aspect that sense decides form, the meaning of all Chinese characters have direct contact with both the word construction and the sentence construction. If the sense representation of the Chinese characters can be realized, the automatical processing of words, phrases and sentences will get a more substantial development.

To represent the original sense of Chinese characters, we must first know the original sense of Chinese characters.

we know the ancient's knowledge to Chinese characters is based on the realization of simple surface characteristic. So, we can explain the sense of Chinese characters by descriptive method. We

elaborate this problem according to the 《The Hand Book Of the original source of common Chinese characters》 edited by Wang Ping:

（1）noun：

【莲】《说文》：芙蕖之实也。　　　　　　　　　　　　　　（means the seed of lotus）

【童】《说文》：男有辠曰奴，奴曰童，女曰妾。从，重省声。　　（means the guilty man ）

（2）verb：

【贺】《说文》：以礼相奉庆也。从贝加声。　　　　　　(means that Send gift to celebrate）

【归】《说文》：女嫁也。从止，从妇省，声。　　　（means that the woman get married）

(3) adjective：

【黄】《说文》：地之色也。从田从，亦声。　　　　　　　　（means the color of soil）

【柔】《说文》：木曲直也。从木矛声。　　　　　　（means the characteristic of wood）

The original sense representation of Chinese characters is to emerge the internal sense hidden in these described language, and appoint them to the input of language. The analytic method of the smallest unit of sense can strip the sense elements off the original sense. What we intend to do is to represent the analyzed senses, and make machines identify them.

3 FOPC and the original sense representation of Chinese characters

When we are discussing the interpretative method in dictionaries, we have already found the correct method to represent the sense of Chinese character--- First Order Predicate Calculus (called FOPC). Based on the logic, FOPC reflects the language world with descriptive language. The most notable character of FOPC is that it only makes few elucidation; The elucidation is very easy to be comprehend; The world that it expresses includes object, the character of object and the relationships with each other.

In syntactic processing, the FOPC approach is flexible, convenient, easily comprehended, and feasible on calculation. The knowledge this method expresses can satisfy our request proposed previously. The well-known program designing language, PROLOG is a language that directly gives order toward computers using FOPC. Compared with the countless sentences in language, the database of Chinese character is really a closed aggregation. If the language PROLOG can deal with all the sentences, then sense of Chinese characters mainly composed of sentences is not beyond the power of PROLOG. In other words, FOPC can provide solid computational foundation to express the original sense of Chinese characters.

Based on the features that people interpret characters, the descriptive method of FOPC, the FOPC's powerful function on language program designing, we contend that FOPC is an efficacious method to represent the original sense of Chinese characters.

FOPC can strengthen the description of events by the method "reification". It can make the event be a measurable object and related to other objects by the previously defined relation. For example:

$w ISA(w, eating)∧Eater(w, Speaker)∧Eaten(w, Apple)

This expression means: there is an event of having a meal; speaker is the doer of the event of eating; apple is the object that is eaten. Here $ is the existential quantifier, ∧ is the logical connective. We call eating is an event, say speaker and apple are the roles of the event.

How to connect the original meaning with FOPC. The key is to find out the components that can act as events from the sense of Chinese characters and the roles of events, and make sure the relation between events and roles. In the examples above, the relation between the event eating and the roles speaker and apple is very simple. We can comprehend it as the relation of eating and eaten.

The ontology theory we will discuss next mainly resolves some problems related with event. Case grammar will resolve the relation between event and roles in sense representation of Chinese characters

4   Ontology and affairs

Ontology is a formal regular description of general conception which can be shared in one domain. This conception is equivalent with the request that we represent the original sense of Chinese characters.From ontology, if Chinese characters sense is abstracted to be topper concepts in the ontology, this topper concept is regarded as a event, and the unit of sense is regarded as the roles of affair, then FOPC can be combined with the original sense of Chinese characters effectively.

Now we represent the examples (except for "贺" and "归") mentioned previously by FOPC on the base of ontology(in this paper the ontology is based on Zhiwei Feng's ONTOL-MT)

【莲】《说文》：芙蕖之实也。

$ w ISA(w, plant)∧owner (w, lotus)∧part (w, seeds)

This expression means: plant is an event, lotus(芙蕖) is one part of plant, seeds is the portion of the lotus. So plant is the attribute of all plant and the topper concept in ONTOL-MT.

$ w ISA (w, people)∧gender (w, male)∧character (w, guilty)∧identity (w, slave)

This expression means: people is an event, the gender of people is male, the feature is guilty, and his identity is a slave. So people is the attribute of all humans and the topper concept in ONTOL-MT.

【黄】《说文》：地之色也。

$ w ISA (w, color)∧color（w, soil）

This expression means: color is an event. Its color is the color of soil. So color is the topper concept in ONTOL-MT.

【柔】《说文》：木曲直也。

$ w ISA(w, character)∧character(w, 能曲能直)∧′s (w, wood)

This expression means: character(feature) is an event, and it describes the feature "能曲能直", and the feature is what wood owns. Feature is the topper concept in ONTOL-MT.

The fact has proved that the concept induced by ontology can be completely integrated with FOPC. speaking Relatively, in ONMTL-MT the words expressing activity involve too many compositions, so it's not easy to express clearly. However, case grammar can resolve this problem for us.

5   Case grammar and FOPC

Describing the sense of Chinese characters with ontology does not resolve the problem with the relation between events and roles, so we define these relations with case grammar.

Some thoughts of case grammar is involved in the analysis of verbs. However, the relation between units is very difficult to define because they are very dispersive and isolated.If we want to combine the sense representation of the verbs with FOPC, we need to improve FOPC on the bases of ontology, combine the representation with case grammar to emerge the relations between cases and events and also the relation of cases with each other. Now we illustrate it with examples(first we should define each relation of cases according to case grammar, but here we don't define them because of the easy examples and limited length of this paper).

【贺】《说文》：以礼相奉庆也。

$ w ISA(w, doings)∧intention (w, celebration)∧mode (w, donate)∧object (w, present)

This expression means: doings is an event. The intention of the event is to celebrate. The mode of celebration is largess, and the donated object is present. So doings is the topper concept in ONTOL-MT. The relation between cases intention and object is combined by mode. They act as the roles of doings.

【归】《说文》：女嫁也。

$ w ISA(w, daily activity)∧mode(w, marry)∧marrier (w, female)∧object (w, male)

This expression means: daily activity is an event. The mode is marriage .The marrier is female. The female married with one male. The relation of cases marrier and object is combined by mode. They compose the roles of doings. So daily activity is the topper concept in ONTOL-MT. The case relation marrier and object is connected with the verb marry. They become the roles of event daily activity.

The predicates, intention, mode, object, etc. are cases in above two examples. To case grammar, many people have calculated the categories of cases of Chinese verbs. Here we need to see that if we extend the concept of case to non-verbs, the results will be more effective. However, there is few study on non-verbs.

Brief summary: now we can see that the way to describe events with FOPC is a logic expression after extracting the units of sense. In brief, the elements that logic conjunctive connects are the units of sense. However, these units share a set of common correlative features, and this set of features are connected by the relation of cases, and congregated in the concepts of ontology. It is the basal method to represent the original sense of Chinese characters.

Reference:
1.　冯志伟，机器翻译研究，中国对外翻译出版公司，2004
2.　沈祖梁、丁利亚，PROLOG语言及其实现，计算机工程，1985，（6）
3.　鲁川、林杏光，动词大词典，中国物资出版社，1994