

Analysis and Processing on the Composing of Noun Conglomeration Combination^①

Liang Xiong & Quan Zhang

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China)

Email :bright84@sina.com

Abstract: The position of noun is replaced by numerous concept types under the system of HNC Theory. Noun is called as a general designation describing objective entity, including type of g、z、r in abstract concepts and p、w、x in concrete concepts. Computers have difficulties in recognizing the relationship of preferential combination in Noun Conglomeration. The paper describes the combination of nouns in different concept types, according as the rule of decreasing differentiation. The rule of concept differentiation sequence among concrete concepts can be summarized as the following: firstly aggregate person concept, secondly individual person concept, thirdly physical concept, the last one is physical property concept. The differentiation degree of physical concept depends on the differentiation degree of semantic network that they hang over.

Keyword: noun; concrete concept; abstract concept; semantic network; HNC; differentiation

1 Introduction

Nouns contribute little to Natural Language Understanding(NLP) when they are used as pure syntax concepts. The position of noun is replaced by numerous concept types under the system of HNC Theory. Noun is called as a general designation describing objective entity, including type of g、z、r in abstract concept and p、w、x in concrete concept.

The definition of noun conglomeration is defined as two or more nouns in succession in a sentence. The Coordinate Combination and Modificatory Combination have obvious language marks, but Noun Conglomeration Combination does not have symbolic mark. Computers have difficulties in recognizing the relationship of preferential combination in Noun Conglomeration. The following will describe the combination of nouns in different concept types, according as the rule of decreasing differentiation. Abstract concepts have their own semantic networks. On the contrary, concrete concepts have to hang over the abstract concepts except a few concept types having their own semantic networks.

2 Between the types of concrete concepts

It is easy to grasp the classification of concrete concept, so the classification of concrete concepts under the system of HNC Theory is exhaustive. HNC divides the concrete concepts into three main bigger types, such as person concept(p), physical concept(w) and physical property concept(x). Then each type is divided into several smaller subclasses. The concrete concept has not especial semantic network except two types—fundamental physical concept(jw) and fundamental physical property concept(jx). For the sake of describing concrete concept, we use the semantic network of abstract concept by the way of hanging over it.

This paper believes it necessary to separate the aggregate person concept from the person concept. As a result, there are four main types in total at last. These four types are just four levels of the degree of concept differentiation. The degree of concept differentiation from upper to the bottom is a gradually decreasing sequence. The degree of concept differentiation inside every level is equal to each other.

First class: The concept of aggregate person, including pe, p- and so so.

Second class: The concept of individual person, including p, pr,rp,pg and so so.

Third class: The concept of physical, including pw, gw, rw, rvw, jw and so so.

^①The work is supported by the National 973 Project of China (#2004CB318104) and the Knowledge Innovation Program of Chinese Academy of Sciences (#13CX04, the Institute of Acoustics) Author Info.: Liang Xiong, born in 1979, male, Ph. D, his research interests include Natural Language Processing; Quan Zhang, born in 1968, male, Ph.D. candidate supervisor.

Fourth class: The nominal concept of physical property, including x, px, gx, jx and so so.

The concept of aggregate person builds on the base of person concept. It is a kind of special concept of person. Since the person has the society attribute, the difference between the aggregate person and the individual person is not only quantity difference any longer. As the saying goes, "Things of one kind come together and birds of a feather flock together." The aggregate person shows many characteristics in concept, especially in the property of socialization. So the differentiation of aggregate person concepts is more than the differentiation of individual person concepts. Since there is the relationship of usage or possession between person concept and physical concept, the socialization property of person concept is more than the socialization property of physical concept. So is the degree of concept differentiation. For example, toy for children(儿童玩具)、feminine belongings(女性用品). "儿童" is in front of "玩具" and "女性" is in front of "用品". The concept of physical property is abstracted on the base of physical concept, so it is on the bottom of concept differentiation tower.

Next comes the detailed analysis of these four levels in turn. The most obvious characteristic of person concepts is that their expressions of concept type includes 'p'. The aggregate person concept is a kind of person concept. Just as its name implies, it is the aggregate of person concepts. The concept of aggregate person includes the following two concept types:

pe: Organization or corporation which is provided with juridical person, such as "部门"(pej40-0), "军队"(pea41).

p-: To express the aggregate of person, such as "百姓", "小组" and so on.

The above two examples of concept types mean organization with multi-person, so named it aggregate person concept. This kind of concept came into being with the development of human civilization and the advance of society. For example, these concepts did not exist during the primitive society. As a result, they have the maximum degree of concept differentiation and they have the highest priority among the concrete concept.

The concept of individual person is the person concept except the concept of aggregate person. Individual person does not mean person concept which can only be singular form. It is for be distinguished from the concept of aggregate person. For example, though "儿童(child)", "少年(youth)" can denote a mass of person, it is still the concept of individual person. We can put forward a simple way to distinguish individual person concept from aggregate person concept.

Decorate the concept using non-condition modifier and see whether the pointing object means a single person. If it means a single person, then it belongs to the concept of individual person, else it belongs to the concept of aggregate person. The combination made up of the concept of individual, such as "one child", "one youth", all mean a kind of person. Because the concept of aggregate person has plural meaning in nature, it holds the natural plural meanings when being decorated by "一个(a, an)". On the contrary, the concept of individual person will come back to singular concept certainly when it is decorated by "一个(a, an)", while it can mean many persons in some certain occasions.

The concept of individual person mainly consists of the concept type of 'p' and some special concept types below.

pr: person with some title of a technical post (such as president, emperor)

rp: person with some honor (such as hero, exemplar and star)

pg: not actual person (such as character in myth or novel)

The most obvious characteristic of physical concept is that 'w' is the mark of physical concept. Physical concept contain mainly the following sorts: man-made physical(pw), common physical(w), fundamental physical(jw). And there are some special physical concept, such as the physical of dynamic effect, the physical of dynamic information, the physical of static state effect. The corresponding example is below this paragraph.

Eg.1: gw: 著作(literature)

rw: 灰烬, 汗水(ash, sweat)

rvw: 寒潮(cold wave)

Among the physical concepts, a majority of concept types have not their own semantic networks except the concept of fundamental physical concept(jw). Most of them have to hang over the semantic network of elementary concept, fundamental concept or synthesis concept. Since the degree of concept differentiation among same level is equal to each other, the concept differentiation depends on the concept differentiation of hanging level. On the sake of processing in a uniform approach, we consider all kinds of physical concepts as hanging structure. The fundamental physical concept, which has its own semantic network, is defined as hanging over the semantic network of concrete concept. As a result, the differentiation of concept depends on the concept differentiation of hanging level, the decreasing sequence is: concrete concept, synthesis concept, elementary concept and fundamental concept. It is just the decreasing differentiation sequence from concrete to abstract.

The above paragraph can be summarized as rule 1 below.

Rule 1: [degree of concept differentiation] Among the physical concepts, the fundamental physical concept, which has its own semantic network, has the maximum degree of concept differentiation. The differentiation degree of physical concepts depends on the differentiation degree of semantic network that they hanging over.

When physical concepts arrange in pairs or groups each other, the fundamental physical concept, which has its own semantic network, has the maximum degree of concept differentiation. The next come the concepts which hang over synthesis concept(pws). The concepts which hang over man-made physical concept(pw) or information physical concept(gw) are situated in the comparatively behind position. According to the opinion of semantic network, the words with concrete concept has the largest concept differentiation, the next comes the synthesis concept between concrete concept and abstract concept. The words that hang over abstract semantic network have the smallest concept differentiation.

Eg.2: 木头家具、稻草房子、猪肉包子、纸杯、塑料杯子

In the above examples, “木头”, “稻草”, “肉” belong to fundamental physical concept, “纸”, “水泥”, “塑料” belong to the concept type of pws, “家具”, “房子”, “包子”, “杯子” belong to the concept type of pw. The concrete concepts also follow the rule of decreasing differentiation sequence: jw>pws>pw.

The most obvious characteristic of physical property concepts is that their expressions of concept type include ‘x’. ‘x’ is the mark of physical property concept. The physical property concepts are divided into two categories of adjectival physical property concepts and nominal physical property concepts. The latter include concept types of xj, xg, xr, xz and so on. They are at the bottom of differentiation level graph of concrete concepts, for reason of it has the least differentiation. Nominal physical property concept is the description of fundamental characteristic of physical and it is often used behind the physical concept. It fits the rule of decreasing differentiation.

Eg 3: 笔记本重量、空气温度

To sum up, the rule of decreasing differentiation among concrete concepts is on the below.

Rule 2: The differentiation of concept type among concrete concept follows the decreasing order of the aggregate person concept, the individual person concept, the physical concept and the nominal physical property concept.

The Statistic to the in real corpors indicates that the arrangement following the rule of decreasing differentiation reach high to 87.3%. In the counterexample of disobeying the order of differentiation, there are two main reasons: The 1st reason is that new words have come into being, as a result, it is not the relationship between semantic chunks. For example, the phrases of positive-sequence, such as “长|兔毛”, “高|鞋跟” belong to the combination of Modificatory Combination firstly and Noun Conglomeration Combination secondly. On the contrary, the example of inverted sequence, such as “金丝猴”, “长颈鹿”, “高跟鞋” belong to the category of new words and these facts negate the rule of decreasing differentiation. The 2nd reason is that the words have multiple properties of concept type.

Eg 4: 飞机驾驶员、汽车司机。

The above examples belong to special examples of person-physical combination. Driver(驾驶员, 司机) belongs to person concept and plane(飞机)、vehicle(汽车) belongs to the concept of man-made physical(pw). According to the rule before, the concept differentiation of person is higher than the one of physical. It seems that these two phrases have disobeyed the rule of sequence that person concept is in the front of physical concept, but it is not true because plane(飞机) and vehicle(汽车) are not only concepts of man-made physical(pw) but also space concepts (pwj2). Space concept belong to the non-condition combination and have the highest priority.

3 Between the types of concrete concepts and abstract concepts

While concrete concept hangs over abstract concept, there is a obvious trend, that is, the higher the differentiation of concept type is, the more the proportion hanging over extending elementary concept is. The lower the differentiation of concept type is, the more the proportion hanging over concept which differentiation of concept type is low is.

From above we have summarized the conclusion Rule 3: Concrete concept with high differentiation of concept type is inclined to hang over the semantic network of abstract with high differentiation.

The statistic result in the knowledge base is shown as table 1 below.

| Type | Hanging extending | Proportion | Hanging element | Proportion | Hanging fundamental | Proportion | Total number | Total Proportion |
|--------------|-------------------|------------|-----------------|------------|---------------------|------------|--------------|------------------|
| AggregateP | 105 | 87.5% | 9 | 7.5% | 6 | 5% | 120 | 100% |
| Individual P | 26 | 21.7% | 57 | 47.5% | 37 | 30.8% | 120 | 100% |
| Physical | 24 | 20% | 63 | 52.5% | 33 | 27.5% | 120 | 100% |

Table 1: the distributing proportion in concrete concepts hanging over semantic networks of abstract concepts

We choose 120 examples of aggregate person concept, individual person concept and physical concept each from word knowledge base of HNC. Then we count the distributing proportion through the semantic networks. From the above table, the difference between the person concept (p) and physical concept (w) in hanging over the three kinds of abstract concepts is not obvious. The reason is that physical concepts contain the concept type of fundamental physical concept (jw) and have the special property of hanging over synthesis concept(s). On statistics of character of physical concept, the numerical value of physical concept in the table get a little increase for these two occasions are not included. Rule 2 is the expansion of Rule 1. The latter one aims at the subclasses inside physical concept, while the former one aims at macroscopical classification of concrete concept.

4 Conclusion

As summing up the above, this dissertation studies the composing of nouns combination according as the rule of decreasing differentiation. The rule of concept differentiation among concrete concepts can be summarized as the following: firstly aggregate person concept, secondly individual person concept, thirdly physical concept, the last one is physical property concept. The differentiation degree of physical concepts depends on the differentiation degree of semantic network that they hang over. The concrete concept with high differentiation of concept type is inclined to hang over the semantic network of abstract concept with high differentiation. The fruits of its labour make up the shortage of SCA and contribute to improving the compositive performance ability of SCA platform.

References

- [1] Huang Zeng-yang. HNC (Hierarchical Network Concept) Theory[M]. Beijing: Tsinghua University Press, 1998.
- [2] Huang Zeng-yang. Mathematics and physics symbol system of language in LCS[M]. Beijing: Ocean Press, 2004.
- [3] Wei Xiang-feng. The software platform for expanded sentence category analysis based on the HNC theory. Beijing: Institute of acoustics, Chinese Academy of Science , doctor's degree paper, 2005.
- [4] Abney, Steven, Parsing by chunks. In Principle_Based Parsing. Berwick, Abney and Tenny (eds.),. Kluwer Academic Publishers,1991.