To Construct the Interpretation Templates for the Chinese Noun Compounds Based on Semantic Classes and Qualia Structures

Xue wei
Dept. of Chinese Lang. & Lit., Peking University/ Research Center of Chinese Linguistics/ Ministry of Education Key Laboratory for Computational Linguistics
Ellen_wx@163.com

Yulin Yuan
Dept. of Chinese Lang. & Lit., Peking University/ Research Center of Chinese Linguistics/ Ministry of Education Key Laboratory for Computational Linguistics
yuanyl@pku.edu.cn

Abstract

This paper focuses on the semantic relations and interpratations of Chinese noun compounds (mostly search terms). In light of the semantic classification from Semantic Knowledge-base of Contemporary Chinese (SKCC) and Qualia Structures introduced by Pustejovsky (1991, 1995), we analyze the combinations of the semantic classes of the noun compounds, and thus, discover the implicit predicates of the noun compounds. Based on these semantic relations of the nouns, we summarize the semantic patterns of the noun compounds and built up an interpretation template database of the paraphrasing verbs for the noun compounds. In conjunction with this database, we further develope an automatic interpreatation program of Chinese noun compounds.

1 Background

As the society is developing rapidly with a lot of new ideas and technologies, noval names sprout out to denote these new concepts, products and etc.. Many noval names are created in the form of noun-noun compound. The phrase pattern "n1+n2" is ambiguous, since it represents different syntactic constructions，such as predication construction, modifier-head construction, appositional construction, and paratactic construction. Among these constructions, the inner semantic relation of the modifier-head "n1+n2" construction is especially complicated. There is a semantic compression with an invisible predicate implied in the noun compounds. Since the predicate is invisible, the semantic relations between the head and the modifier are not quite clear[1].For example, "木头桌子"(the wood table, which means that the table is made of wood), "爱情故事"(love story, which means that the story is about love[2]), "钢材仓库"(steel warehouse, which has two different meanings: the warehouse to store

[1] Some modifier-head constructions have implicit nouns, for example, "封面女郎"(cover girl, which means that the girl whose photos are on the cover) .
[2] Cf. Yuan Yulin (1995).

26th Pacific Asia Conference on Language,Information and Computation pages 609–619

steel, and the warehouse made of steel[3]). We suggest that the implicit predicate could be the paraphrase verb that reveals the semantic relations between the modifier and the head in noun compounds. Thus, the aim of this paper is to discover the implicit predicate of the noun compounds and to generate the paraphrases of the noun compounds.

The modifier-head noun compounds are basic constructions in almost all languages. As they are "derivative, easily composed but ambiguious"(Wang Meng, et al. 2010), they have aroused much interest in Theoretical Linguistics and Computational Linguistics. As Wang Meng et al. (2010) points out, the research on Chinese noun compounds interpretation can be applied in the fields like question answering, information retrieval and lexicography. We suggest that the noun compounds interpretation is crucial in information retrieval.

A basic information retrieval process contains the following steps: submiting searching request → sending the request → sorting → seaching index → selecting pages → ranking results → presentation of the results. The information retrieval appears to be a simple behavior accomplised in just a few seconds, while a lot of analysis and operations are needed after a simple query[4]. The operating procedures in information retrieval are generally divided into two parts: one part is to analyze the users' search intention which is top-down, and the other part is to analyze the structure and meaning of the searching words which is bottom-up. Both parts are important to obtain the required results.

Therefore, if we want to interprate the noun compounds automatically, we need to understand the ontological meaning of the noun compounds that submitted by the web user, and provide references of the users' search intentions as well. For example, when a user inputs "蔬菜大王"(vegetable king) as the searching word, we guess that he maybe wants to know the news about "蔬菜大王"(vegetable king). However, the noun compound "蔬菜大王"(vegetable king) happens to be an ambiguous noun compound. It might means someone who sells/buys vegetables, or someone who plants vegetables, or someone who eats vegetables. If we can decote these different meanings of the compound "蔬菜大王", we can provide the different searching results for the user. Thus, to recover the implicit predicates of the noun compounds is helpful to understand the users' search intentions.

In order to have a better understanding of the search intentions, we have collected 850 Chinese noun compounds from the daily top search terms of Baidu news[5] (2010.9 to 2011.4) and some other literature texts. Besides the basic analysis of the semantic relations and the implicit predicates of these noun compounds, we need the following steps to arrive at their semantic patterns: (1) we summarize the combination patterns according to the semantic classes of the nouns from SKCC. We thus can predict the implicit predicates according to the semantic classes of the modifier and head nouns; (2) in light of the Qualia Structures introduced by Pustejovsky (1991, 1995), we find out that most implicit verbs of the noun compounds are agentive roles or telic roles of the head noun. We thus treat them as paraphrase verbs to reveal the semantic relations of the noun compounds. (3) In the base of the paraphrase verbs, we build up a paraphrase database and an

[3] Cf. Zhou Ren (2007).

[4] Cf. Sina Reports on Science and technology, Mar. 12, 2012. http://www.sina.com.cn.

[5] http://top.baidu.com

automatic paraphrase program of the noun compounds.

2 The semantic classification of nouns

The Semantic Knowledge-base of Contemporary Chinese (SKCC) is a large scale Chinese semantic resource developed by the Institute of Computational Linguistics of Peking University. It provides a large amount of semantic information such as semantic hierarchy and collocation features for 66,539 Chinese words and their English counterparts (Wang and Yu, 2003). Because the classification of nouns is designed for the need of grammatical research (Wang Hui, et al. 2006) and is based on grammatical analysis (Wang and Yu, 2003), we adopt this classification standard as the basis to construct the interpretation templates of the noun compounds. The semantic classification of nouns in SKCC is as follows:

1 thing
1.1 entity
1.1.1 organism
1.1.1.1 person
1.1.1.1.1 individual
1.1.1.1.1.1 name
1.1.1.1.1.2 profession
1.1.1.1.1.3 identity
1.1.1.1.1.4 relation
1.1.1.1.2 group
1.1.1.1.2.1 organization
1.1.1.1.2.2 society
1.1.1.2animal
1.1.1.2.1 beast
1.1.1.2.2 bird
1.1.1.2.3 insect
1.1.1.2.4 fish
1.1.1.2.5 reptile
1.1.1.3 plant
1.1.1.3.1 tree
1.1.1.3.2 grass
1.1.1.3.3 flower
1.1.1.3.4 crop
1.1.1.4 microbe
1.1.2 object
1.1.2.1 artifact
1.1.2.1.1 building
1.1.2.1.2 clothes
1.1.2.1.3 food
1.1.2.1.4 drug
1.1.2.1.5 cosmetics
1.1.2.1.6 works
1.1.2.1.7 software
1.1.2.1.8 hardware
1.1.2.1.9 asset
1.1.2.1.10 bill
1.1.2.1.11 certificate
1.1.2.1.12 symbol
1.1.2.1.13 material
1.1.2.1.14 instrument
1.1.2.1.14.1 tool
1.1.2.1.14.2 vehicle
1.1.2.1.14.3 weapon
1.1.2.1.14.4 furniture
1.1.2.1.14.5 musical-instrument
1.1.2.1.14.6 electricity
1.1.2.1.14.7 stationery
1.1.2.1.14.8 sports-instrument
1.1.2.2 natural object
1.1.2.2.1 celestial body
1.1.2.2.2 geography
1.1.2.2.2.1 land
1.1.2.2.2.2 water
1.1.2.2.3 weather
1.1.2.2.4 mineral
1.1.2.2.5 element
1.1.2.2.6 substance
1.1.2.3 excrement
1.1.2.4 shape
1.1.3 part
1.1.3.1 body-part
1.1.3.2 object-part
1.2 abstraction

1.2.1 attribute
1.2.1.1 measurable
1.2.1.2 fuzzy attribute
1.2.1.2.1 property_of_human
1.2.1.2.2 description_of_event
1.2.1.2.3 property_of_object
1.2.1.3 color
1.2.2 information
1.2.3 field
1.2.4 rule
1.2.5 physiological_state
1.2.6 psycho feature
1.2.6.1 feelings
1.2.6.2 cognition
1.2.7 motivation
2 process
2.1 event
2.2 natural phenomenon
2.2.1 visible phenomenon
2.2.2 audible phenomenon
3 space
3.1 location
3.2 direction
4 time
4.1 specific time
4.2 relative time

3 The Qualia Structures of nouns: Agentive roles and Telic roles

The generative lexicon theory (here after GLT), which is proposed by Pustejovsky (1991, 1995), has a great impact in the field of linguistics and natural language processing. Based on the computation and cognition background, this theory deals with natural language semantics, in particular the semantics of words, both alone and in combination, i.e. the problem of compositionality. It aims to explain the meanings of words in the specific contexts by using a detailed description of semantics of words and building a limited semantic operation mechanism.

GLT has divided the semantics of words into four levels: argument structure, qualia structure, event structure and lexical inheritance structure. Argument structure is a specification of number and type of logical arguments, and how they are realized syntactically. Qualia structure is the modes of explanation that includes Formal, Constitutive, Telic and Agentive roles. Event structure is the definition of the event type of a lexical item and a phrase, whose sorts include State, Process, and Transition. Lexical inheritance structure is the identification of how a lexical structure is related to other structures in the type lattice, and its contribution to the global organization of a lexicon. A set of generative devices connects these four levels, providing for the compositional interpretation of words in context (Pustejovsky, 1995: 61).

The qualia structure is inspired by Aristotel's *Four Causes*. A qualia structure has four roles: constitutive role is the relation between an object and its constituents, or proper parts (including Material, Weight, Parts and Component elements); formal role is the basic category which distinguishes the object within a larger domain (including Orientation, Magnitude, Shape, Dimensionality and so on); telic role is the purpose and function of the object; agentive role is the factors involed in the origin or "bringing about" of an object. In fact, a noun's qualia structure illustrates the things, events and relationships related to the object, which is very helpful to the interpretation of noun compounds.

In light of this idea, we find that most implicit predicates of the noun compounds (n1+n2) are n1 or n2's telic roles or agentive roles. So we can use nouns' telic roles or agentive roles to build the database of

interpretation templates of noun compounds[6].

4 The cognitive basis of noun compounds

From the perspective of cognition, every noun compounds (n1+n2) has a hidden event (we call it "background event"). When events in concept are expressed in the level of language，it always includes verbs and the arguments dominated by the verbs. The words of n1 and n2 are usually the arguments of the verbs. For example, "红木家具"(mahogany furniture), whose background event is making furniture by mahogany. "Mahogany" is the Material role of "make", while "furniture" is the Product role of "make". Another example is "体操奶奶"(gymnastics grandma). Its background event is that a grandma does gymnastics. "Grandma" is the Agent role of "do", while "gymnastics" is the Result role of "do".

When the speaker wants to emphasize a certain semantic role (noun) of the event in a declarative way, *de* structure , a correspondent analytic pattern of NN compound, such as "NP1+V+的+NP2" or "V+NP1+的+NP2", could also be used. In Chinese, particle *de* is ususuall considered as a marker introducing a relative clause for the head.

When the speaker use *de* structure, the the head of the noun is emphasized, while the modifiers，namely the verb and other arguments， are downgraded in the *de* structure, for example, "（用）红木制作的家具"(the furniture which is made of mahogany) and "做体操的奶奶"(the grandma who does gymnastics). When the speaker usese "n1+n2"pattern, he wants to emphasize both the head n2 and the modifier n1, while the verb connecting n1 and n2 is omitted in the phonological level. For example, we use the Material "红木"(mahogany) to be the modifier, the Product "家具"(furniture) to be the head, and get the noun compound"红木家具"(mahogany furniture). Another example is that we use the Activity "体操"(gymnastics) to be the modifier and the Agent "奶奶"(grandma) to be the head, and then we get the noun compound "体操奶奶"(gymnastics grandma).

The listener usually intends to decode the noun compounds in the background events, which is built on the basis of common sense. Therefore, we can interpret the noun compounds just in a reverse process. We need to recover the semantic roles of n1 and n2 through their semantic classes and find out the predicate that dominates them. Then we can recover the whole background event completely. Especially, to find the verb that dominates the two nouns is the key to interpret the noun compound (n1+n2)[7].

As different kinds of noun compounds have different kinds of background events and implicit verbs, we have summarized different interpretation templates that express different background events from 850 noun compounds instances (n1+n2). Among these interpretation templates, we find that most implicit verbs of the noun compounds (n1+n2) are n1 or n2's telic roles or agentive roles. For example, the explanation of "摩托妈妈"(Motorcycle Mom) is "骑/坐/造/修摩托的妈妈"("the mom who rides on/makes/repairs the motorcycle"). In this case, the semantic class pattern is "artifact+ relation". The verb "骑/坐"(ride on) is the telic role of n1"摩托"(motorcycle), and the verb "造/

[6] Song Zuoyan (2010) has pointed out that the implicit predicate of noun compounds could be gotten by n1 or n2's telic roles or agentive roles. But the details need to be further investigated and generalized.

[7] Cf. Yuan Yulin (1995).

修”(make/repair) is the agentive role of n1“摩托”(motorcycle). Another example is “司机餐馆”(drivers restaurant), whose explanation is “（专门供）司机吃饭的餐馆”(the restaurant is specially for the drivers). The semantic class pattern is “occupation+ building”. The verb “吃饭”(eat) is the telic role of n2 “餐馆”(restaurant). In the interpretation templates, we indicate the roles of the verbs. Meanwhile, we add their telic roles and agentive roles in the noun knowledge database. We thus build up a data model that is based on the knowledge and approach of linguistics for the interpretation of noun compounds.

5 The computation procedures and the interaction between semantic patterns and interpretation templates

5.1 The computation procedures Electronically-available resources

Based on the analysis in the above chapters, we deal with the 850 instances of the noun compounds (n1+n2) in the following proeduress:

⑴ Use the segmentation software to split all the noun compounds (n1+n2) into n1+n2.

⑵ Find all n1s’ and n2s’ semantic classes in SKCC and describe the semantic class combination patterns with the semantic classes of n1 and n2. We abstract the tokens of noun compounds into types of combination patterns. We thus can predict the implicit predicates (paraphrasing verbs) according to the semantic classes of the modifier and head nouns

Since the lexicon database in SKCC is limited, we can add an unknown word’s semantic class manually.

⑶ Paraphrase the interpretation template with implicit predicates for every noun compound. We also specify the roles of the verbs. Is it the role of n1 or n2, and is it an agentive role or telic role? If it is a qualia structure role of n1, we mark it as v1; if it is a qualia structure role of n2, we mark it as v2.

⑷ Every noun compound (n1+n2) has a semantic class combination pattern and an interpretation template. We sort out these semantic class patterns and interpretation templates to build up a noun-noun coordination database.

5.2 The interaction between semantic patterns and interpretation templates

We have summarized 326 semantic class patterns (here after semantic patterns) and 208 interpretation templates in total. These semantic patterns can be divided into two classes: ⑴ a sematic pattern in correspondence with one interpretation template; ⑵ a semantic pattern in correspondence with two or more interpretation templates.

⑴ a sematic pattern in correspondence with one interpretation template;

We have gotten 212 such semantic patterns, and 62 corresponding interpretation templates. We choose ten interpretation templates and the corresponding semantic patterns randomly, and list them below:

i. If the semantic class of n1 is “tool” and the semantic class of n2 is “cognition”, the interpretation template is “(通过)+n1+表现+的+n2” ((through)+n1+express+De+n2). The verb “表现” (express) can be seen as the agentive role of n2. For example, “瓷器爱国主义” (china patriotism), the interpretation is “(通过)瓷器表现的爱国主义”(the patriotism which is expressed through the china).

ii. If the semantic class of n1 is

"relative time" and the semantic class of n2 is "field", the interpretation template is "n1+产生+的+n2/产生于+n1+的+n2" (n1+be produced+De+n2/Be produced in+n1+De+n2). N1 is the time when n2 is/was produced. For example, "当代文学" (contemporary literature), the interpretation is "当代产生的文学/产生于当代的文学" (the literature which is produced in contemporary age).

iii. If the semantic class of n1 is "organization" and the semantic class of n2 is "location", the interpretation template is "n1+建立+的+n2" (n1+build+De+n2). The verb "建立" (build) can be seen as the agentive role of n2. For example, "网易养猪场" (Wangyi Pig farm), the interpretation is "网易建立的养猪场" (The pig farm which is built by Wangyi).

iv. If the semantic class of n1 is "profession" and the semantic class of n2 is "organization", the interpretation template is "供+n1+v2+的+n2" (For+n1+v2+De+n2). The verb v2 is the telic role of n2. For example, "民工学校" (migrant workers school), the interpretation is "供民工读书/上学的学校" (the school for migrant workers to study).

v. If the semantic class of n1 is "physiological_state" and the semantic class of n2 is "microbe", the interpretation template is "引起+n1+的+n2" (cause+n1+De+n2). For example, "流感病毒" (flu virus), the interpretation is "引起流感的病毒" (viruses that cause flu).

vi. If the semantic pattern is "field+event [8]", or "property_of_object+abstraction", or "property_of_object+artifact", the interpretation template is "是+n1+(性+)的+n2" (is+n1+De+n2). The corresponding examples are "历史机遇" (historical opportunity), "基础项目" (basic project), "基础设施" (infrastructure construction), their corresponding interpretations are "是历史(性)的机遇" (the opportunity which is historical), "是基础(性)的项目" (the project which is basic), "是基础(性)的设施" (the installation which is basic).

vii. If the semantic pattern is "tool+artifact", or "profession+society", or "profession+group", the interpretation template is "由+n1+构成+的+n2" (by+n1+constitute+De+n2). The corresponding examples are "电脑网络" (computer network), "工人阶级" (worker class), "义工组织" (volunteer organization), their corresponding interpretations are "由电脑构成的网络" (the network which is constituted by computers), "由工人构成的阶级" (the class which is constitued by workers), "由义工构成的组织" (the organization which is constituted by volunteers).

viii. If the semantic pattern is "name+relation[9]", or "name+feelings", the interpretation template is "n1+拥有+的+n2" (n1+own+De+n2). The corresponding examples are "汪峰女儿" (WangFeng's daughter), "梁咏琪恋情" (LiangYongqi's love affair), their corresponding interpretations are "汪峰拥有的女儿" (the daughter who is owned by WangFeng), "梁咏琪拥有的恋情" (The love affair which is owned by LiangYongqi).

ix. If the semantic pattern is "building+material", or "food+drug", the interpretation template is "v1+n1+用+的+n2" (v1+n1+use+De+n2). The verb v1 is the agentive role of n1 (such as "修建" (build), "制作" (make), etc.). The corresponding examples are "建筑钢材" (building steel), "食品添加剂" (food additives), their corresponding

[8] It means that n1's semantic class is field and n2's semantic class is event. Follwings are the same.

[9] Most nouns that are n2s are monovalent nouns, and n1 is an argument of n2.

interpretations are “修建建筑用的钢材” (the steels that are used for building), “制作食品用的添加剂” (the additives that are used for food).

x. If the semantic pattern is “location+cosmetics”, or “location+excrement”, or “location+food”, or “location+tool”, the interpretation template is “产自+n1+的+n2” (produce in+n1+De+n2). N1 is n2’s place of origin. The corresponding examples are “法国香水” (French perfume), “南海珍珠” (South Sea pearls), “信阳毛尖” (Xinyang tea), “景德镇瓷器” (Jingdezhen china), their corresponding interpretations are “产自法国的香水” (the perfume which is produced in French), “产自南海的珍珠” (the pearls which are produced in South Sea), “产自信阳的毛尖” (the tea which is produced in Xinyang), “产自景德镇的瓷器” (the china which is produced in Jingdezhen).

⑵ a semantic pattern in correspondence with two or more interpretation templates.

We divided this situation into two types: ① a semantic pattern has two interpretation templates; ② a semantic pattern has three or more interpretation templates.

① a semantic pattern has two interpretation templates

We have collected 88 semantic patterns of this type, and 100 corresponding interpretation templates. We choose four interpretation templates and the corresponding semantic patterns randomly, and list them below:

i. If the semantic class of n1 is “event”, and the semantic class of n2 is “location”, the two interpretation templates are: a. “发生+n1+的+n2” (happen+n1+De+n2); b. “有+n1+的+n2” (have+n1+De+n2). N2 is the place where n1 happens. For example, “交通路口” (traffic crossing), the corresponding interpretations are: “a.发生交通的路口; b.有交通的路口” (the crossing where traffic happens).

ii. If the semantic class of n1 is “drug”, and the semantic class of n2 is “animal”, the two interpretation templates are: a. “喂了+n1+的+n2” (feed+n1+De+n2); b. “吃了+n1+的+n2” (eat+n1+De+n2)[10]. For example, “瘦肉精羊” (drug sheep), the corresponding interpretations are: “a.喂了瘦肉精的羊 (the sheep which is fed with drugs); b.吃了瘦肉精的羊 (the sheep which eats drugs)”.

iii. If the semantic pattern is “name+works”, or “name+event”, or “identity+works”, and the two interpretation templates are: a. “n1+v2+的+n2” (n1+v2+De+n2), the verb v2 is the agentive role of n2 (such as “发表” (publish),“表演” (perform),“写” (write), etc.); b. “关于+n1+的+n2” (about+n1+De+n2). The corresponding examples are “鲁尼声明” (Rooney statement), “刘谦新魔术” (LiuQian new magic), “小学生日记” (primary school student diary), and their corresponding interpretations are: “a. 鲁尼发表的声明 (the statement which is published by Rooney), b. 关于鲁尼的声明 (the statement about Rooney)”; “a. 刘谦表演的新魔术 (the magic which is performed by Liu Qian), b. 关于刘谦的新魔术 (the magic about LiuQian)”; “a. 小学生写的日记 (the diary which is written by a primary school student), b. 关于小学生的日记 (the diary ahout a primary school student)”.

iv. If the semantic pattern is “organization+society[11]”, or

[10] Animals won’t take the initiative to eat medicine or additives, so if the sematic class of n1 is “drug”, and the semantic class of n2 is “animal”, the relation between n1 and n2 is not the initative to eat, but the passive feeding. Through the entailment: X feed Y Z → Y eat Z, “喂了+n1+的+n2”(feed +n1+ De+n2) can entail “吃了+n1+的+n2”(eat +n1 +De +n2). Cf. Yuan Yulin and Wang Minghua (2009, 2010).

[11] Organization is usually founded by people, and has certain social functions. So all members in the organization have the character “work”, and belong to

"organization+identity", or "group+ society", the two interpretation templates are: a. "在+n1+工作+的+n2" (in+n1+work+De+n2); b. "属于+n1+的+n2" (belong+n1+De+n2). The corresponding examples are "企业员工" (enterprise employee), "委员会成员" (committee members), "消防队人员" (fire brigade staff), their corresponding interpretations are: "a. 在企业工作的员工 (the employees who work in the company), b. 属于企业的员工 (the employees who belong to the company)"; "a. 在委员会工作的成员 (the members who work in the committee), b. 属于委员会的成员 (the members who belong to the committee)"; "a. 在消防队工作的人员 (the staff who work in the fire brigade), b. 属于消防队的人员 (the staff who belong to the fire brigade)".

②**a semantic pattern has three or more interpretation templates**

We have gotten 26 semantic patterns of this type, and 46 corresponding interpretation templates. We choose four interpretation templates and the corresponding semantic patterns randomly, and list them below:

i. If the semantic class of n1 is "organization", and the semantic class of n2 is "abstraction", and the three interpretation templates are: a. "n1+v2+的+n2" (n1+v2+De+n2), the verb v2 is the agentive role of n2 (such as "创造" (create), "设计" (design), etc.); b. "n1+拥有+的+n2" (n1+own+De+n2); c. "供+n1+使用+的+n2" (for+n1+use+De+n2). For example, "国家财政" (state finance), the corresponding interpretations are: "a. 国家制定的财政 (the finance which is formulated by state); b. 国家拥有的财政 (the finance which is owned by state); c. 供国家使用的财政 (the finance which is used by state)".

ii. If the semantic class of n1 is "food", and the semantic class of n2 is "event", the three interpretation templates[12] are: a. "v1+n1+的+n2" (v1+n1+De+n2), the verb v1 is the telic role of n1 (such as "吃"(eat), etc.); b. "n1+引起+的+n2" (n1+cause+De+n2); c. "关于+n1+的+n2" (about+n1+De+n2). For example, "兴奋剂事件" (dope event). The corresponding interpretations are: "a. 吃兴奋剂的事件 (the event which is taking dope); b. 兴奋剂引起的事件 (the event which is caused by dope); c. 关于兴奋剂的事件 (the event which is about dope)".

iii. If the semantic pattern is "location+profession", or "space+profession", the three interpretation templates are: a. "来自+n1+的+n2" (come from+n1+De+n2); b. "在+n1+v2+的+n2" (in+n1+v2+De+n2), the verb v2 is the telic role of n2 (such as "教书" (teach), etc.); c."在+n1+工作+的+n2" (in+n1+work+De+n2). N1 can be the place where n2 comes from, or the place where n2 works. The corresponding examples are "上海工人" (Shanghai workers), "中学教师" (middle school teachers), their corresponding interpretations are "a. 来自上海的工人 (the workers who come from Shanghai), b. 在上海上班的工人 (the workers who work in Shanghai), c. 在上海工作的工人 (the workers who work in Shanghai)"; "a. 来自中学的教师 (the teachers who come from middle school), b. 在中学教书的教师 (the teachers who teach in the middle school), c. 在中学工作的教师 (the teachers who work in the middle school)".

iv. If the semantic class of n1 is

the organization.

[12] These three templates express the semantic information of noun compounds from detailed or enriched to less. In the first template, we know the detail of the event by the telic role of n1; in the second template, we only know that the event is caused by n1, but don't know how it is caused; in the third template, we only know that the event is related to n1, but don't know how it is related.

"field" and the semantic class of n2 is "abstraction", the six interpretation templates are: a. "关于+n1+的+n2" (about+n1+De+n2), n1 is the content of n2. For example, "法律常识" (law commonsense). Its corresponding interpretation is "关于法律的常识" (the commonsense which is about law); b. "在+n1+领域/方面内+存在+的+n2" (in+n1+field+exist+De+n2). For example, "政治把柄" (politics handle). Its corresponding interpretation is "在政治领域/方面内存在的把柄" (the handle which exists in politics); c. "v2+n1+的+n2" (v2+n1+De+n2), the verb v2 is the telic role of n2 (such as "经营"(operate), etc.). For example, "化工行业" (chemistry industry). Its corresponding interpretation is "经营化工的行业" (the industry which is engaged in chemistry); d. "n1+（上）+使用+的+n2" (n1+(top)+use+De+n2). For example, "工业技术" (industry technology). Its corresponding interpretation is "工业上使用的技术" (the technology which is used in industry); e. "由+n1+组成+的+n2" (by+n1+compose+De+n2). For example, "社会环境" (society environment). Its corresponding interpretation is "由社会等诸因素组成的环境" (the environment which is composed of society and other factors); f. "考虑+n1+的+n2" (consider+n1+De+n2). For example, "政治头脑" (politics mind). Its corresponding interpretation is "考虑政治（方面问题）的头脑" (the mind which considers about politics). There are many meanings of this semantic pattern. That's because the meanings of nouns of filed and nouns of abstraction are very fuzzy, and it's hard to decide which verb connects the two nouns.

6 Summary

All in all, we build up an interpretation template database that contains the paraphrasing verbs for noun compounds in Chinese. Based on this database, we exploit some other language resources (such as Hownet) to generate the telic roles and agentive roles of every noun automatically. Finally, we develop a program to automatically interpret the meanings of the noun compounds. The accuracy of this program is 94.23% by manual evaluation. [13]

References

Rumshishy, Anna and James Pustejosky. 2011. Generative Lexicon Theory: Theoretical and Empirical Foundations. PPT of *Summer School* at the Center for Chinese Linguistics, Peking University.

Dong, Zhendong and Dong Qiang. *Hownet*. Website: http://www.keenage.com.

Ungerer, Friedrich and Hans-Jörg Schmid. 2006. *An Introduction to Cognitive Linguisitics—Second Edition*. Pearson Education Limited.

Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics*, 17(4): 209–441.

Pustejovsky, James. 1995. *The Generative Lexicon*, Cambridge, Massachusetts: The MIT Press.

Li, Xiaoming, Yan Hongfei and Wang Jimin. 2005. *Search Enging—Principle, Technology and System*. Beijing: Science Press.

Song, Zuoyan. 2009. *Research on the Event Coercion in Mandarin Chinese*. Ph.D. thesis, Peking University.

Song, Zuoyan. 2010. On the Affixoid Which Can Trigger Off Event Coercion. *Chinese Teaching in the World*, 24(4): 446-458.

[13] Cf. Wei Xue (2012) §7.2: The result of test and analysis.

Tan, Jingchun. 2010. Semantic relations between nouns and noun modifiers and their roles in dictionary definition. *Chinese Language*, 337: 342-355.

Wang, Hui and Yu, Shiwen. 2003. The Semantic Knowledge-base of Contemporary Chinese and its Applications in WSD. In the *Proceedings of the Second SIGHAN workshop on Chinese Language Processing*, ACL.

Wang, Hui, Yu Shiwen and Zhan Weidong. 2003. New Progress of the Semantic Knowledge-base of Contemporary Chinese(SKCC). *Language Computing and Content-based Text Processing*. Beijing: Tsinghua University Press.

Wang, Hui, Zhan Weidong and Yu Shiwen. 2003. The Specification of the Semantic Knowledge-base of Contemporary Chinese. *Journal of Chinese Language and Computing*, 13 (2): 159-176.

Wang, Hui, Zhan Weidong and Yu Shiwen. 2*006.* Structure and Application of the Semantic Knowledge-base of Modern Chinese. *Applied Linguistics*, 1: 134-141.

Wang, Meng, Huang Chu-ren, Yu Shiwen and Li Bin. 2010. Chinese Noun Compound Interpretation Based on Paraphrasing Verbs. *Journal of Chinese Information Processing*, 24(6): 3-9.

Wang, Meng. 2010. *Linguistic Knowledge Acquisition of Noun for the Construction of Probablistic Lexical Knowledge-base*. Ph.D. thesis, Peking University.

Wei, Xue. 2012. *Research on Chinese Noun Compound Interpretation for Semantic-Query*. M.A. thesis, Peking University.

Yuan Yulin. 1995. On the Implicit Predicate and its Syntactic Consequences. *Chinese Language*. 4: 241-255.

Yuan, Yulin. 2008a. A Programme of Semantic Resources Oriented to Information Retrieval. *Linguistic Sciences*, 7(1): 1-11.

Yuan, Yulin. 2008b. *Cognitive-based Studies on Chinese Computational Linguistics* (Essays). Beijing: Peking University Press.

Yuan, Yulin and Wang Minghua. 2009. The Types of Textual Entailment and their Inference Mechanisms. *Chinese Linguistics*, 3: 123-138. Beijing1: Peking University Press.

Yuan, Yulin and Wang Minghua. 2010. The Inference and Identification Models for Textual Entailment. *Journal of Chinese Information Processing*, 24(2): 3-13.

Zhang, Xiusong and Zhang Ailing. 2009. An Introduction to the Generative Lexicon. *Contemporary Linguistics*, 3: 267-271.

Zhou, Ren. 2007. The Principle of Information Amount and the Rhythm Mode of Syntactic Combination in Mandarin Chinese, *Chinese Language*, 3: 208-222.