A Study on the Taxonomy of Chinese Noun Compounds

Lulu Wang[1(✉)] and Meng Wang[2]

[1] Communication University of China, Beijing 100024, China
`lulu.wang@cuc.edu.cn`
[2] School of Humanity, Jiangnan University, Wuxi 214122, China
`wangmengly@163.com`

Abstract. This paper deals with the semantic interpretation of Chinese noun compounds. We first propose a novel taxonomy of Chinese noun compounds based on the syntagmatic possibility and semantic transparency. Further, we analyze the qualia structure of the nouns and their relations with the hidden verbs. In the end, we propose a fine-grained classification and the interpretation patterns of the noun compounds.

Keywords: Noun compound · Semantic relation · Qualia roles

1   Introduction

A Noun Compound[1] syntactically behaves as a single noun, but is semantically interpreted as a compressed proposition, which could be elaborated by the hidden verb (Yuan 1995). For example, "爱情故事" (love story) could be illustrated as "讲述爱情的故事" (a story that tells about love), in which the verb "讲述"（to tell） infers the missing relation between "爱情" (love) and "故事" (story). Many researchers (Tan 2010, Zhou 2010, Wang 2010 and Wei 2013) suggest that the hidden verbs are crucial to the interpretation of noun compounds and design various ways to reveal the hidden verbs. However, some noun compounds could not be reconstructed by the hidden verbs, such as (4a) below.

(1)   a.机组　成员　（crew　member）

　　　b.机组　的　成员[2]　（crew　PART　member）

　　　c.构成　机组　的　成员　（constitute　crew　PART　member）

(2)   a.钻石　戒指　（diamond　ring）

　　　b.?钻石　的　戒指　（diamond　PART　ring）

　　　c.镶嵌　钻石　的　戒指　（inlay　diamond　PART　ring）

---

[1]   We only analyze the noun compounds (N1-N2) that N1 behaves as the modifier of N2.

[2]   "的" is an attribute particle in Chinese and here it is shorted for as PART.

(3) a.试管　婴儿　（test tube　baby）

　　　b.*试管　的　婴儿　（test　tube　PART　baby）

　　　c.?试管　培育　的　婴儿　（test　tube fertilize　PART　baby）

(4) a.夫妻　肺片　（the　spouse　pork　lungs）

　　　b.*夫妻　的　肺片　（the　spouse　PART　pork　lungs）

　　　c.*夫妻　制作　的　肺片　（the　spouse　make　PART　pork lungs）

In the above examples, only (1a) allows the attribute particle "的" inserted in between the noun compounds, while others are not grammatical as (2b-4b) show. Notably, the noun compounds in (1c-3c) could be interpreted by the hidden verbs. However, (4c) is not acceptable with the hidden verb "制作" (make), this is because "夫妻肺片" means a famous Sichuan style dish of pork lungs in chili source, which entails an idiomatical meaning which could not be reconstructed compositionally. Thus, we suggest that the semantic interpretation of noun compounds varies in the semantic transparency: some expressions are literal, some are inferable by verbs and some are idiomatic.

In this paper, we focus on the semantic interpretation on Chinese noun compounds. We will suggest a novel taxonomy of Chinese noun compounds based on the transparency of the compounds. Further, we will have a deep understanding on the qualia structure of the nouns. A fine-grained classification of the noun compounds is provided in the end.

2   Previous Studies

2.1   Noun Compound Semantics

In the theoretical field, linguists describe the semantics of the noun compounds via a set of abstract relations, as represented in the work of Levi (1978) and Warren (1978). Following this tradition, some researchers in the computational field focus on the taxonomies of noun compounds. Ó Séaghdha (2007) proposed six semantic relations: BE, HAVE, IN, ACTOR, INST, ABOUT, and each relation is subdivided into subcategories. For example, HAVE is subdivided into the subcategories of possession, condition-experiencer, property-object, part-whole and group-member. Tratz and Hovy (2010) presented a large, fine-grained taxonomy of 43 noun compound relations, which are notably tested by the Amazon's Mechanical Turk service. However, the approach of abstract rela-tions is problematic in several ways. As Nakov and Hearst (2013) pointed out that (1) it is unclear which relation inventory is best, (2) relations capture only part of the semantics, and (3) multiple relations are possible. For example, Wei (2012: 40) assumes that "中国电影" (Chinese movies) is classified into the categories of LOCATION and CONTENT.

Considering these drawbacks, other researchers use verbal paraphrases to interpret the noun compounds. Finn (1980) interpreted "salt water" with dissolved in. Butnariu and Veale (2008) summarized 8 relational possibilities. For example, "headache pill" might be paraphrased as "headache-inducing pill", "headache prevention pill", "pill for treating headaches", "pill that causes headaches", "pill that is prescribed for headache" and "pill that prevents headaches". With these verbs, the paraphrases are more specific than that of the abstract relations.

Accordingly, there are two ways to interpret the noun compounds semantics in Chinese. Wang (2010) adopts a bottom-up strategy to capture the verbs of noun compounds and provides four types of paraphrase patterns. As Wei (2012) pointed out, these four types are not specific enough to give proper interpretations. Instead, Wei classifies the noun compounds into 8 major types and 346 subcategories, which are proved to be fine-grained. However, some of these subcategories can be merged, and some noun compounds belong to more than one subcategory. Further, they all neglected the semantic transparency of the noun compounds, which should be taken serious consideration.

2.2   Semantic Transparencies

Recalling the above example of (4a), the noun compound "夫妻肺片" is not decomposable, that is, the meaning of the compound is not simply the combinations of the literal meanings of the parts. Levi (1978) argues for a transparency scale for the noun compounds as in the following table.

Table 1. Levi's transparency scale of noun compounds

	types	examples
a	transparent	orange peel
b	partly opaque	grammar school
c	exocentric	ladybird
d	partly idiomatic	flea market
e	completely idiomatic	honeymoon

In the table above, Levi summarizes five types of noun compounds based on the semantic transparency. These different types show the different interpretation patterns of the noun compounds. For example, "orange peel" is simply the combination of the parts of "orange" and "peel". But "grammar school" cannot be combined literally, that is because a hidden verb should be revealed to illustrate this compound as in "grammar teaching school". In contrast, the other types cannot be combined literally, nor be interpreted by the hidden verb. For instance, "ladybird" is not a kind of bird, but a kind of bug "Coccinellidae"[3]. And "honeymoon" has nothing to do with "honey" or "moon", but refers to the vacation that brides and grooms celebrate their marriage. The Type of "partly idiomatic" is special, because it is partly idiomatic that verbs are not easy to recover. It is not acceptable to say "flea selling market", but means market selling small commodities. Enlightened by Levi's ideas, we will present a novel taxonomy of Chinese noun compounds based on the semantic transparency.

---

[3]   Here "lady" is named after "Virgin Mary", see more: http://www.hkhk.edu.ee/nature/ladybird_legends.html

3   Taxonomy of Chinese Noun Compounds

In light of Levi's transparence scale and Nun-berg, Sag and Wasow (1994)'s claim on idioms, we collected 428 noun-noun compounds (N1-N2) and classified them into the following four categories.

Table 2. Basic types of noun compounds

	transparency scale	examples
a	transparent	机组人员 (crew member)
b	partly opaque	钻石戒指 (diamond ring)
c	partly idiomatic	试管婴儿 (test tube baby)
d	completely idiomatic	夫妻肺片 (the spouse pork lung)

As the table shows, the first three types are decomposable at the syntagmatic level, but the last one is non-decomposable. We suggest that non-decomposable idioms should be analyzed as a whole unit both syntactically and semantically. The other types are decomposable, which can be divided into N1 and N2. However, the semantic relations of these types are different in semantic transparency. We will illustrate these differences with the following examples.

(5)   a.机组　人员　（crew-member　*crew members* ）
　　　　b.印尼　火山　（ Indonesia-volcano　*an Indonesia volcano* ）

The examples in (5) belong to type a, which are decomposable and the semantic relations are transparent. Syntactically, N1 is the attribute of N2. The attribute particle "的" can be inserted in between N1 and N2. Semantically, the meaning of the compound is the combination of the literal meaning of the parts. We suggest that this type is the default type of the noun-noun pairs in Chinese, where N2 is the head noun and N1 is the modifier of the noun. The semantic relations are possessive, locative and time.

(6)   a.钻石　戒指　（diamond-ring　*diamond rings* ）
　　　　b.围棋　高手　（chess-master　*chess masters* ）

The examples in (6) are decomposable as well, but the attribute particle cannot be inserted in between. If it is inserted in between, it sounds odd to native speakers. That is because they are different from type 'a' that there are hidden verbs in between the nouns. For example, "钻石" (diamond) and"戒指" (ring) imply the verb of "镶嵌" (inlay); "围棋" (chess) and "高手" (master) imply the verb of "下". Hence, these expressions are partly opaque that we need to reveal the hidden verbs to interpret these noun compounds. In the next section, we will further explore the subcategories of this type within the qualia structure theory.

(7)   a.试管　婴儿　　（test tube-baby　*test tube babies*）
        b.皮包　公司　　（suitcase-company　*paper company*）

Similar to type b, the examples in (7) are also decomposable, but the semantic relations are more complex than the former one. In (7a), "试管" (test tube) and "婴儿" (baby) cannot be combined literally, that is because "试管" (test tube) metaphors as in vitro (glass) fertilization. Such compressed concept cannot be simply revealed by the lexical semantics, such as "装" (fill in), but can be revealed by the verbs co-occurred in the text, such as "培育" (fertilize). As for (7b), the case is the same in that we cannot analyze the compound by the literal meanings of "皮包" (suitcase), but comprehend them as an idiomatic expression. Hence, such type is the most complex in semantic relations that they are not simply the combinations of the literal meanings of the parts, but involve a process of metaphors, which enhance the difficulty in revealing the hidden verbs.

4   Revealing the Hidden Verbs

In light of Aristotle's arguments on the four causes, Pustejovsky (1995) proposed four qualia roles of nouns: formal, constitutive, agentive and telic. Wei (2012) first adopts this idea into the interpretation of Chinese noun compounds and discovers some relations between the qualia structure and the noun compounds interpretations. Further, Yuan (2014) analyzes the qualia structure of nouns in Chinese and proposes some new qualia roles based the data of Chinese, such as action and handle. To take "食品" (food) for example, the qualia roles of food are listed in the following table.

Table 3. The qualia roles of "食品" (*food*)

	食品	food
formal	商品	commodity
constitutive	快餐	fast food
unit	篮子	basket
evaluation	新鲜	fresh
agentive	制作	make
telic	吃	eat
handle	购买	buy

We believe that the above qualia roles are more fine-grained to Chinese data, and thus we analyze the qualia structure of nouns based on Yuan's qualia structure knowledge system.

We analyzed 428 noun compounds on their semantic relations and the qualia roles of the Head noun. It appears to us that there is a clear correspondent relationship between the semantic relations and the qualia roles of the head noun. To illustrate, we summarize this correspondence in the following table.

Table 4. The semantic relations of noun compounds

	semantic relations	qualia roles	interpretation patterns	examples
1	posses-sive	CON	"属于" (be-long)+N1+de+N2	机组人员 (crew member)
2	property	EVA/ TEL	"性质是"(the feature is) +N1+de+N2	可行性报告 (operative report)
3	locative	CON/ AGE	"位于" (locate) +N1+de+N2; "在" (at) +N1+V+de+N2	雅典奥运会 (Athens Olympics)
4	time	CON/ AGE	"在" (in)+N1+de+N2	梅雨季节 (plum rain season)
5	content	CON/ TEL	"关于" (about) +N1+de+N2; V+N1+de+N2	爱情故事 (love story)
6	material	AGE	"用" (with) +N1+V+de+N2	钻石戒指 (diamond ring)
7	patient	TEL	V+N1+de+N2	围棋高手 (chess master)
8	actor	ACT	N1+V+de+N2	教委文件 (the board of edu-cation document)
9	cause	AGE	N1+V+de+N2	考试焦虑 (test anxiety)
10	partly-idiomatic	-	metaphoric or metonymic meaning of N1+de+N2	试管婴儿 (test tube baby)
11	idiomatic	-	idiom	夫妻肺片 (the spouse pork lungs)

In the table above, we have summarized 10 subcategories of noun compounds based on their semantic relation. To interpret these noun compounds, we present different interpretation patterns with different conditions. The first category of "posses-sive" in the above table corresponds to the noun compounds of type 'a' in table 2, which could be interpreted as the literal meanings of the parts. To illustrate, the paraphrasing sentence of "机组人员" (crew members) is "属于机组的人员" (the members that belonged to the crew).

From (2) to (5), these four types correspond to both type 'a' and type 'b', since the meaning of the compound could not only be interpreted by the fixed pattern of the components, but also could be predicted by the hidden verb. For instance, the paraphrasing verb of the compound "雅典奥运会" (Athens Olympics) could be "举办" (to hold), and thus the paraphrasing sentence will be "在雅典举办的奥运会" (The Olympic Games that was held in Athens.). As for "爱情故事" (love story), it could be paraphrased as "关于爱情的故事" (the story about love), and also "讲述爱情的故事" (the story telling about love).

Moreover, categories from (6) to (8) correspond to type 'b', in which the hidden verb must be revealed. In this group, the qualia roles of the head noun are different in each type. For example, the qualia role in (6) is AGE, which is because material usually relates to the MAKE relation. And the relation of "patient" in (7) relates more with TELIC roles, which are interpreted as the functions of N1. For example, "围棋高手" (chess master) could be paraphrased as "下围棋的高手" (the masters of playing chess). Here, "下（棋）" (to play) is the TELIC role of "围棋" (chess).

The last two types correspond to type c and the non-decomposable idioms separately. Noun compounds as in (10) should be interpreted with the metaphoric meaning, and thus they cannot be interpreted with the hidden verbs. To take "试管婴儿" (test tube babies) as an example, we cannot illustrate the compound using the expressions like "在试管里孕育的婴儿" (the babies that are fertilized in test tubes). The word "试管" (test tubes) has the metonymic meaning as a "试管孕育技术" (in glass fertilization). Therefore, we need to infer the metaphoric meaning of such compounds as "用试管技术孕育的婴儿" (the babies that are fertilized by the technic using test tubes). For the idioms, they are not decomposable at all and should be treated as a whole unit. For example, "夫妻肺片" only refers to the name of a dish.

5   Concluding Remarks

This paper deals with the semantic interpretation of Chinese noun compounds. We first propose a novel taxonomy of Chinese noun compounds based on the transparency of the compounds. Further, we analyze the qualia structure of the nouns and propose a fine-grained classification of the noun compounds. In the future, we will implement this taxonomy of noun compounds in some practical NLP tasks, such as semantic similarity computation, question answering system and search query engines.

References

1. Butnariu, C., Veale, T.: A concept-centered approach to noun-compound interpretation. In: Proceedings of the 22nd International Conference on Computational Linguistics, pp. 81–88 (2008)
2. Finn, T.: The semantic interpretation of compound nominal. Ph.D. Dissertation, University of Illinois, Urbana (1980)
3. Levi, J.: The Syntax and Semantics of Complex Nominals. Academic Press, New York (1978)
4. Nakov, P.I., Hearst, M.A.: Semantic Interpretation of Noun Compounds using Verbal and Other Paraphrases. ACM Trans. Speech Lang. Process. **10**(3), 13 (2013)
5. Nunberg, G., Sag, I.A., Wasow, T.: Idioms. Language **70**(3), 491–538 (1994)

6. Séaghdha, D.O.: Designing and evaluating a semantic annotation scheme for compound nouns. In: Proceedings of the 4th Corpus Linguistics Conference (2007)
7. Pustejovsky, J.: The Generative Lexicon. The MIT Press (1995)
8. Tan, J.C.: Semantic Relations between nouns and noun modifiers and their roles in dictionary definition. Studies of Chinese Language **4** (2010). (in Chinese)
9. Tratz, S., Hovy, E.: A taxonomy, dataset, and classifier for automatic noun compound interpretation. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 678–687 (2010)
10. Wang, M.: Linguistic Knowledge Acquisition of Noun for the Construction of Probablistic Lexical Knowledge-base. Ph.D. Dissertation, Peking University (2010). (in Chinese)
11. Warren, B.: Semantic patterns of noun-noun compounds. In: Gothenburg studies in English 41, Acta Universitatis Gothoburgensis, Goteburg (1978)
12. Wei, X.: Research on Chinese Noun Compound Interpretation for Semantic-Query. Master Thesis. Peking University (2012). (in Chinese)
13. Yuan, Y.L.: Weici Yinhan Jiqi Jufa Houguo: de zi jiegou de chengdai guize he de de yufa, yuyigongneng. Studies of Chinese Language **4** (1995). (in Chinese)
14. Yuan, Y.L.: A Study of Chinese Semantic Knowledge System Based on the Theory of Generative Lexicon and Argument Structure. Journal of Chinese Information Procession **27**(6), 23–30 (2013). (in Chinese)
15. Zhou, R.A.: Studies on the syntax and semantics of noun-noun compounds. China Social Science Press (2010). (in Chinese)