Annotating Qualia Relations and Types in Chinese Compound nouns

Song Zuoyan and Zhao Qingqing

School of Chinese Language and Literature,
Beijing Normal University, Beijing, China, 100875
meszy@163.com; 874540721@qq.com

ABSTRACT. *This paper introduces a semantic markup lexicon of Chinese compound nouns based on Generative Lexicon theory (GL), especially the qualia structure and types proposed by this theory. A case study of compounds containing hua shows that qualia structure and types play an important role in Chinese word formation, which can help refine the analysis of Chinese compounds. Moreover, this study reveals that the head noun can indicate qualia information as well as the modifier. Also, some productive compounding patterns are identified. Different from previous studies, this study is also concerned with qualia relations between the compound and the elements.*

Keywords: compound nouns, annotation, Generative Lexicon, qualia structure, types

1. **Introduction.** Compounding, which is more productive in analytic languages such as Chinese, has been proven to be a pervasive means of word formation in many languages. Among all the compounding constructions, nominal compound construction is the most productive and many new words are compound nouns. The analysis of compound nouns has always been regarded as a hard task for natural language processing (NLP).

The qualia structure in the Generative Lexicon theory (GL) proposed by Pustejovsky [1] has been introduced into the analysis of the semantic relations in nominal compounds by Johnston and Busa and sheds much light on this issue. Based on data from both English and Italian, they point out that the qualia structures of the nouns in a compound provide relational structure enabling compositional interpretation of the modification of the head noun by the modifying noun [2].

Recently, the approach of qualia modification has been adopted in several research works to analyze Chinese compound nouns. Lee et al. discuss the qualia modification in noun-noun compounds found in Chinese as well as a couple of other languages like German, Spanish, Japanese and Italian, demonstrating that the proposed qualia structure enables compositional interpretation within the compound [3]. Wang and Huang examine the adjectival modification to nouns especially in two cases: "*chǎng+* noun" and "adjective+ *túshuguǎ*" [4]. Also, they specifically investigate the modifier-head type in

compound event nouns [5]. Song and Qiu examine the qualia relations in the nominal compounds containing verbal elements (NCCVs) and identify some productive compounding patterns. Different from previous studies which focus mostly on the qualia relation between two elements in a compound, this work aims to reveal what qualia information of compound nouns can be indicated by the verb or the verbal morpheme(s) , and how it is indicated [6].

This paper aims to introduce an annotation scheme for Chinese compound nouns based on GL, especially the qualia structure and types proposed by this theory. The rest of this paper is structured as follows. Section 2 provides a brief overview to the qualia structure and types in GL. Section 3 describes the annotation framework and methods. Section 4 gives a case study of compounds containing *hua* (花). Finally, we summarize the paper in Section 5.

2. **Qualia Structure and   Types in GL.** In GL, Qualia structure contains four basic qualia roles [1]:

- Constitutive role: The relation between an object and its constituents or proper parts.
    e.g. material, weight, parts and component elements;
- Formal role: That which distinguishes the object within a large domain.
   e.g. shape, color, orientation, magnitude, dimensionality and position;
- Telic role: purpose and function of the object.
    i.e. purpose that an agent has in performing the object and built-in function or aim which specifies certain activities.
- Agentive role: factors involved in the origin or "bringing about" of an object.
   e.g. creator, artifact, natural kind and causal chain.

The representation of qualia structure is as follows:

$$
\begin{bmatrix}
\alpha \\
QUALIA = \begin{bmatrix}
CONST = \text{what x is made of} \\
FORMAL = \text{what x is} \\
TELIC = \text{function of x} \\
AGENTIVE = \text{how x came into being}
\end{bmatrix}
\end{bmatrix}
$$

FIG. 1. THE REPRESENTATION OF QUALIA STRUCTURE

The qualia are taken as representing an essential component of word meaning, capturing how language speakers understand objects and relations in the world and providing the minimal explanation for the linguistic behavior of lexical items. For example, for the noun *book*, *text* is a constitutive role noun, *red* is a formal role adjective, *read* is a telic role verb and *write* is an agentive role verb.

According to Pustejovsky[7-8], the domain of individuals can be structured into three, increasingly complex, conceptual types:

- Natural types: Natural kind concepts making reference only to Formal and Constitutive qualia roles;
  e.g. rock, water, woman, tiger and tree.
- Artifactual types: Concepts making reference to Telic (purpose or function), or Agentive (origin).
  e.g. knife, beer, husband and dancer.
- Complex types: Concepts making reference to a relation between at least two types from the other levels.
  e.g. book, lunch, university and temperature

Similarly, the domains of relations and properties are also partitioned into three ranks as shown in (1) and (2).

(1) N: fall, walk, rain, put, have
    A: donate, spoil, quench
    C: read, perform

(2) N: red, large, flat
    A: useful, good, effective
    C: rising, frightened

3. **Annotation Framework and Methodology.** Unlike the annotation scheme proposed for nominal compounds by Pustejovsky et al. [9] and Bouillon et al.[10], which annotate compounds in context, we annotate compounds in a lexicon. Moreover, we annotate types as well as qualia relations in a compound.

In addition to the noun-noun construction, we examine other nominal compounding constructions at play, such as verb-noun, verb-verb, noun-verb and so on. Before tagging the semantic information, we build a lexicon of Chinese nominal compounds by extracting data from Hownet Lexicon (version 2000). Besides, many compounds are collected by annotators. Specifically, the following information of a compound will be provided:

1. Each element's part-of-speech (P) in the compound, e.g. adjective (a), verb (v) and noun (n).
2. The types of the compound and each element (T), i.e. natural type (N), artifactual type (A) and underspecified type (U).
3. The structural relation between two elements in the compound, e.g. modification construction, subject-verb construction and verb-object construction.
4. The qualia relation between two elements in a compound or the qualia relation between the compound and the elements, i.e. Formal, Constitutive, Telic and Agentive.
5. Sense variations. e.g. metaphors and metonymies.

Some examples are given in table 1.

Taking advantage of this lexicon, we can make quantitative and qualitative analysis of the compound nouns. In the next section, we will present a case study of compounds containing *hua*.

TABLE 1. A SEMANTIC MARKUP LEXICON OF CHINESE COMPOUND NOUNS BASED ON GL

W_c	W_e	P_1	P_2	construction	T_1	T_2	T_NC	Relation
砂纸	sand paper	n	n	modification	N	A	A	Constitutive
白纸	white paper	a	n	modification	N	A	A	Formal: color
废纸	waste paper	a	n	modification	A	A	A	Telic
墙纸	wall paper	n	n	modification	A	A	A	Telic
打印纸	printing paper	v	n	modification	A	A	A	Telic
竹纸	paper made from young bamboo	n	n	modification	A	A	A	Agentive

4. A Case Study of Compounds Containing *hua(花)*

4.1. **Overview.** The noun *hua* has several senses which have close relations with each other. We will confine our study to the following three senses with high frequency:

－ 1) hua_1: the part of a plant which is often brightly colored, grows at the end of a stem, and only survives for a short time.
－ 2) hua_2: small plants that are grown for their flowers as opposed to trees, shrubs, and vegetables   .
－ 3) hua_3: anything resembling a flower

Among these senses, *hua₁* is the basic sense, *hua₂* is a meaning derived from *hua₁* by the means of metonymy and *hua₃* is a meaning metaphorically derived from *hua₁*. 204 compounds containing *hua* are collected and the majority of them are modification constructions. Table 2 shows that *hua* tends to serve as the head noun of a compound rather than the modifier, but the distribution of frequencies associated with each sense is very uneven. *Hua₁* can be both modifier and head noun. *Hua₂* prefers modifier and *hua₃* is likely to be head noun.

TABLE 2. THE DSTRIBUTION OF FREQUNCIES ASSOCIATED WITH EACH SENSE OF *HUA*

hua-X:70;34%			X-*hua*:134;68%		
hua₁-X	hua₂-X	hua₃-X	X-hua₁	X-hua₂	X-hua₃
42; 21%	26; 13%	2; 1%	91; 45%	1; 0%	42;21%

4.2. **X-hua.**

4.2.1. **X-*hua₁*.** Generally speaking, only three kinds of qualia relations can be seen in X-*hua₁* compounds, i.e. Constitutive, Formal and Telic, among which Formal is the most frequent, closely followed by Constitutive. It is reasonable since *hua₁* is a natural type noun which usually makes reference only to formal and constitutive qualia roles.

Table 4 presents some examples. The compound *tao-hua₁* shows a constitutive relation since the head noun *hua₁*(blossom) is a part of the modifier noun *tao* 'peach (tree)'. *Gan-hua₁*, *la-ba-hua₁* and *ji-guan-hua₁* all represent the formal relation. Both *la-ba* and *ji-guan* are used metaphorically. The former specifies the shape of the head noun *hua₁* and the latter specifies the color as well as the shape of *hua₁*. In the compound of *jiu-hua₁*, the modifier *jiu* describe the purpose of *hua₁*, which is used to brew beer.

TABLE 3. QUALIA RELATIONS IN X-*HUA* COMPOUNDS

	Formal	Constitutive	Agentive	Telic	total
X-hua$_1$	*45；49%*	*39；43%*	0；0%	7；8%	91
X-hua$_2$	1；100%	0；0%	0；0%	0；0%	1
X-hua$_3$	1；2%	*28；67%*	5；12%	8；19%	42

TABLE 4

桃花	tao-hua$_1$	peach(tree)-blossom	'peach blossom'	Constitutive: whole_part
玫瑰花	mei-gui-hua$_1$	rose-flower	'rose'	Constitutive: whole_part
干花	gan-hua$_1$	dry-flower	'dry flower'	Formal
喇叭花	la-ba-hua$_1$	trumpet-flower	'morning glory'	Formal: shape
鸡冠花	ji-guan-hua$_1$	cockscomb-flower	'cockscomb'	Formal: Shape, color
酒花	jiu-hua$_1$	beer-flower	'hops'	Telic

Tao-hua$_1$ represents the most common pattern in constitutive relations: the head nouns *hua$_1$* are used to specify a subpart of the denotation of the modifier, namely a kind of plant. Similar examples are *huai-hua* 'sophora japonica', *li-hua* 'pear flower' and *xing-hua* 'apricot blossom'. Although *mei-gui-hua* is also an example of this pattern, it is different from *tao-hua* in that it is a complex type which can refer to both a flower as shown in (3) and a plant as shown in (4).

(3) 我　摘　　　　了　　　一　　朵　　玫瑰花。
　　wo　zhai　　　le　　　yi　　duo　mei-gui-hua
　　I　　pluck　ASP　one　CL　　rose
　　'I plucked a rose.'

(4) 我　　种　　　　了　　一　　棵　玫瑰花。
　　wo　zhong　　le　　yi　　ke　mei-gui-hua
　　I　　grow　　ASP　one　CL　rose
　　'I grew a rose.'

In other words, *mei-gui-hua* is used metonymically. A word that denotes flowers is used to refer to the plant with these flowers. This difference comes from the fact that a rose is cultivated as an ornamental for its flowers while a peach tree (*tao-shu*) is usually cultivated for its fruits. Similar examples are *he-hua* 'lotus', *ju-hua* 'chrysanthemum' and *hai-tang-hua* 'begonia'. In fact, apart from constitutive relation, this compounding patter also presents agentive relation, for a flower is produced by a plant.

4.2.1 **X-*hua$_3$***.Table 3 shows that all the four qualia roles can be seen in X-hua3 compounds, among which, Constitutive relation is the most frequent. Examples are given in table 5.

TABLE 5. X-*HUA₃*

绢花	juan-hua₃	silk-flower 'silk flower'	Constitutive
纸花	zhi-hua₃	paper-flower 'paper flower'	Constitutive
麻花	ma-hua₃	fibre-flower 'fried dough twist'	Formal
浴花	yu-hua₃	bath-flower 'cockscomb'	Telic
发花	fa-hua₃	hair-flower 'hair flower'	Telic
雕花	diao-hua₃	carve-flower 'carving'	Agentive

Note that unlike the constitutive relation in X-*hua₁* compounds, which presents a *whole_part* relation (*hua₁* specifies a subpart of the denotation of the modifier noun), in these compounds, the modifier specifies the material of which the head noun is composed. There are two productive patterns.

1) N+*hua₃*=N: Both the modifier and the compound are natural types. As pointed out by Lee et al. (2010), the modifier noun serves the main semantic content and the head noun *hua₃* specifies the shape of the modifier noun. For example, *shui-hua₃* water spray' is not a kind of flower but a kind of water in the shape of a flower. More examples are given in table 6.

TABLE 6. EXAMPLES OF PATTREN 1: N+*HUA₃*=N

水花	shui-hua₃	water-flower	'water spray'	Constitutive
雪花	xue-hua₃	snow-flower	'snowflake'	Constitutive
火花	huo-hua₃	fire-flower	'spark'	Constitutive
霜花	shuang-hua₃	frost-flower	'frost'	Constitutive
雨花	yu-hua₃	rain-flower	'rain flower'	Constitutive
冰花	bing-hua₃	ice-flower	'frost/rime'	Constitutive
泪花	lei-hua₃	tear-flower	'tear'	Constitutive
浪花	lang-hua₃	wave-flower	'spindrift/spray/waves'	Constitutive
葱花	cong-hua₃	onion-flower	'chopped green onion'	Constitutive
油花	you-hua₃	oil-flower	'blob of oil'	Constitutive
盐花	yan-hua₃	salt-flower	'a little salt/a pinch of salt'	Constitutive

2) A/N+*hua₃*=A: In this pattern, the compounds are artifactual type, regardless of the types of the modifiers. Unlike compounds like *shui-hua₃*, which are natural types and have nothing to do with intention, these compounds presents a made_of relation and therefore are artifactual types. For instance, *zhi-hua₃* 'paper flower' is a fake flower made of paper rather than a real kind of flower. Here, *hua₃* is interpreted as the shape of an object made of paper. To put it in another way, both the constitutive quale and formal quale (i.e. shape) of

this compound can be seen, which is indicated by the modifier and head noun respectively. In these compounds, the modifiers do not serve the main semantic content as seen in the compounds like *shui-hua₃*. For example, *zhi-hua₃* is not a kind of paper in the shape of a flower, just as *dou-hua₃* is not a kind of bean. Table 7 provides more examples.

TABLE 7. EXAMPLES OF PATTREN 1: A/N+*HUA₃*=A

绢花	juan-hua₃	silk-flower	'silk flower'	Constitutive
蜡花	la-hua₃	candle-flower	'candle flower'	Constitutive
萝卜花	luobo-hua₃	radish-flower	'fower made of radish'	Constitutive
豆花	dou-hua₃	bean-flower	'tofu pudding'	Constitutive
蛋花	dan-hua₃	egg-flower	'egg flake'	Constitutive

What is common to these two groups of examples above is that the object that is denoted by the head noun is understood to be an image of that object, i.e. not an instance of the object itself. This shows that the qualia of the compound not only can be activated by the modifier but also by the head noun.

4.3. **hua-X.** Table 8 shows that constitutive relation and telic relation are the most frequent in *hua₁*-X compounds and *hua₂*-X respectively. Some examples are given in Table 9. The most common constitutive relation is also *whole_part* relation as seen in compounds like *tao-hua₁.* Exactly speaking, it is not the head noun but the compound that specifies a subpart of the modifier *hua₁*. For example, a corolla is a part of a flower. Similar examples are *hua-bao* 'bud', *hua-jing* 'stem' and *hua-rui* 'stamen'.

TABLE 8. QUALIA RELATIONS IN X-HUA COMPOUNDS

	Constitutive	Formal	Agentive	Telic	total
hua₁-X	*31;74%*	6;14%	3;7%	2;5%	42
hua₂-X	5;19%	2;8%	2;8%	*17;65%*	26
hua3-X	0	2	0	0	2

TABLE 9. *HUA*-X

花冠	hua₁-guan	flower-hat	'corolla'	Constitutive
花名	hua₃-ming	flower-name	'name of a flower'	Formal
花瓶	hua₁-ping	flower-vase	'flower vase/vase'	Telic
花蜜	hua₁-mi	flower-honey	'nectar'	Agentive
花丛	hua₂-cong	flower-cluster	'flower in cluster/flowering shrub'	Constitutive
花秧	hua₂-yang	flower-seeding	'flower seedling'	Formal
花铲	hua₂-chan	flower-scoop	'flower scoop'	Telic
花籽	hua₂-zi	flower-seed	'flower seeds'	Agentive
花卷（儿）	hua₃-juan	flower-roll	'steamed twisted roll'	Formal

It is interesting that both *cai-hua₁* and *hua₁-cai* mean 'cauliflower', but they have different qualia relations, i.e. TELIC and COSTITUTIVE respectively. In the former compound, the modifier *cai* represents the purpose of the head noun *hua₁*, which is used as a kind of vegetable. In the latter, the modifier *hua₁* specifies the material of the head noun *cai*.

TABLE 10. *CAI-HUA₁ VS. HUA₁-CAI*

菜花	cai-hua₁	vegetable-flower	'cauliflower'	Telic
花菜	hua₁-cai	flower-vegetable	'cauliflower'	Constitutive

4.4. **Summary.** In summary, the most frequent two qualia relations in compounds containing *hua* are FORMAL and CONSTITUTIVE. In other words, *hua* is a natural type whose formal role (especially shape) and constitutive role are salient. On one hand, the formal quale of *hua* is often modified as seen in the compound *la-ba-hua*. On the other hand, when *hua* serves as a head noun, its formal role (i.e. shape) is also activated, whereby its basic sense is extended metaphorically to refer to something resembling a flower. It can be inferred from the constitutive relation that a flower, which is a part of a plant, has several parts.

It turns out that some compound nouns present two different qualia relations because the head nouns *hua* also indicate qualia information. Additionally, qualia relations can be seen not only between two elements in a compound but also between the compound and the elements.

5. **Conclusion.** This paper introduces a semantic markup lexicon of Chinese compound nouns based on Generative Lexicon theory (GL), especially the qualia structure and types proposed by this theory. A case study of compounds containing *hua* shows that qualia structure and types play an important role in Chinese word formation, which can refine the analysis of Chinese compounds. Moreover, this study reveals that the head noun can indicate qualia information as well as the modifier. Also, some productive compounding patterns are identified. Different from previous studies, this study is also concerned with qualia relations between the compound and the elements in it.

Up to now, about 4,000 compounds have been annotated. For future work, more compounds need to be examined. Furthermore, we hope that the extracted generalizations such as productive compounding patterns and their interpretation templates can be used in NLP, such as unknown words recognition and their automatic interpretation. On the other hand, we expect to finally build an ontology of Chinese words based on the lexicon.

REFERENCES

[1] Pustejovsky, J.: Generative Lexicon. MIT Press, Cambridge (1995)

[2] Johnston, M., Busa, F.: Qualia structure and the compositional interpretation of compounds. In: E. Viegas (ed). Breadth and Depth of Semantics Lexicons, pp. 67~187. Dordrecht, Kluwer (1999)

[3] Lee, C., Chang, C., Hsu, W., Hsieh, S.: Qualia Modification in Noun-Noun Compounds: A Cross-Language Survey. In: Proceedings of the 22nd Conference on Computational Linguistics and Speech Processing (ROCLING-2010), pp.379~390.(2010)

[4] Wang, S., Huang, C.: Adjectival Modification to Nouns in Mandarin Chinese: Case Studies on "cháng+ noun" and "adjective+ túshūguan". In: Proceedings of Pacific Asia Conference on Language, Information and Computation. Tohoku University, Sendai, Japan(2010)

[5] Wang, S., Huang, C.: Compound Event Nouns of the 'Modifier-head' Type in Mandarin Chinese. In: Proceedings of the 54th PacificAsia Conference on Language, Information,and Computation. Nanyang Technological University, Taiwan(2011)

[6] Song, Z., Qiu, L.: Qualia Relations in Chinese Nominal Compounds Containing Verbal Elements. International Journal of Knowledge and Language Processing. 4(1), 1-15(2013)

[7] Pustejovsky, J.: Type Construction and the Logic of Concepts, In: Bouillon, P., Busa, F. (eds.) The Syntax of Word Meanings, pp.91-123. Cambridge University Press, Cambridge (2001)

[8] Pustejovsky, J.: Type Theory and Lexical Decomposition. Journal of Cognitive Science.6, 39–76 (2006)

[9] Pustejovsky, J., Rumshisky, A., Moszkowicz, J.L., Batiukova, O.: GLML: A Generative Lexicon Markup Language. In: Proceedings of the Generative Lexicon Workshop. Instituto di Linguistica Computazionale (CNR), Pisa, Italy, September (2008)

[10] Bouillon, P., Jezek, E., Melloni, C., Picton, A.: Annotating qualia relations in Italian and French complex nominal. In: ański, Piotr et.al.(eds), Proceedings of the LREC-2012 Workshop on "Challenges in the Management of   Large Corpora" (CMLC).pp.1527-1532.(2012)