International Journal of Knowledge and Language Processing
Volume 4, Number 1, **2013**

www.ijklp.org
KLP International ©2013 ISSN 2191-2734
pp.**1–15**

Qualia Relations in Chinese Nominal Compounds Containing Verbal Elements

Song Zuoyan[1] and Qiu Likun [2]
[1] School of Chinese Language and Literature,
Beijing Normal University, 100875, Beijing, China
meszy@163.com

[2] School of Chinese Language and Literature,
Ludong University, 264025, Yantai, China
qiulikun@pku.edu.cn

Received January 2013; revised March 2013

ABSTRACT. *From the perspective of qualia structure, this article addresses the semantic relations of a special class of nominal compounds in Chinese which contains verbal elements. It turns out that the verbal elements in different constructions can indicate the telic role or agentive role of the nominal compounds in different ways. Nearly all the nominal compounds are artifactual types and most of the head nouns are artifactual types too. The VN compound construction in which the V indicates the telic or agentive role is highly productive and some compounding patterns are identified. By contrast, TELIC-selecting patterns are more than AGENTIVE-selecting patterns. In addition, this paper explores the differences between Chinese and English in word formation. The analysis of qualia information in NCCVs presented in this paper could be potentially useful for a variety of NLP tasks such as natural language processing (especially disambiguation), automatic interpretation of unknown words and Chinese-English machine translation.*

Keywords: qualia structure; Generative Lexicon Theory; Telic; Agentive; referential transparency; disambiguation

1. **Introduction.** Nominal compounds have been proven to be a pervasive construction in many languages. The analysis of nominal compound constructions is a recalcitrant problem for linguistic semantics and poses serious challenges for natural language processing systems. A great deal of work has been done to capture the structure and the semantic of nominal compounds. In recent years, the qualia structure in the Generative Lexicon Theory (GLT) has been introduced into the analysis of the semantic relations in nominal compounds and shed much light on this issue [1,3,5,11]. However, most researches focus on the noun-noun compounds in various languages and no enough attention has been paid on other compound constructions. In fact, in addition to the noun-noun construction, there

are other nominal compounding constructions at play, such as verb-noun, verb-verb, noun-verb and so on [4]. Although these constructions are not as productive as noun-noun construction, they also represent important compounding process especially in analytic languages such as Chinese. Therefore, they should receive much more attention.

This paper aims to investigate the qualia relations in the nominal compounds containing verbal elements (NCCVs). There is a great deal of disagreement over the definition of compound in Chinese. We agree on Li and Thompson's opinion, considering as compounds all polysyllabic units that have certain properties of single words that can be analyzed into two or more meaningful elements, or morphemes, even if these morphemes cannot occur independently in modern Chinese [4]. Different from previous studies on noun-noun compounds, which focusing on the qualia relation between the two nouns, this work aims to reveal what qualia information the verb or the verbal morpheme(s) in the NCCVs can indicate and how, and further find out the most productive compounding patterns. For example, in the compound *wo-shi* (literally, lie-room) 'bed room', the verbal morpheme of *wo*'lie' indicates the purpose of the compound *wo-shi*, which is for lying (sleeping). In other words, *wo* is the telic role of the compound. In GL lexical entry, events are encoded in the TELIC and the AGENTIVE roles, so our analysis will focus on these two roles. The data in this study are from Hownet dictionary (2000).

The rest of this paper is organized as follows. Section 2 provides a brief overview to the qualia structure and type system in GLT. Section 3 describes the qualia relations in three types of NCCVs. Section 4 gives a list of productive compounding patterns in NCCVs. Section 5 explores the difference between Chinese and English in word formation based on the analysis of NCCVs. Section 6 discusses some applications of this study in natural language process. Finally, we summarize the paper in Section 7.

2. **Qualia structure and type system in the Generative Lexicon Theory.** In GLT, Pustejovsky [7, 8] proposes to describe lexical expressions by using four representation levels, argument structure, event structure, qualia structure, and lexical typing structure. The qualia structure of a given noun incorporates (at most) the following four roles:

a. FORMAL: the basic category that distinguishes the object within a larger domain. e.g. orientation, magnitude, shape, dimensionality, color and position.

b. CONSTITUTIVE: the relation between an object and its constituents, or proper parts. e.g. materials, weight, parts and component elements.

c. TELIC: the purpose or function of the object, if there is one; used-for or functions-as relation.

d. AGENTIVE: the factors involved in the object's origins or "coming into being"; created-by relation. e.g. creator, Natural kind, Artifact or causal chain.

The qualia structure can be seen as providing functional tags to words, linking the words to a network of concepts as shown in (1).

$$(1)\begin{bmatrix}\alpha \\ \text{ARGSTR} : \begin{bmatrix}\text{ARG1} : x \\ \dots\end{bmatrix} \\ \text{EVSTR} : \begin{bmatrix}\text{EV1} : e_1 \\ \dots\end{bmatrix} \\ \text{QUALIA} : \begin{bmatrix}\text{CONST} : \text{what } x \text{ is made of} \\ \text{FORMAL} : \text{what } x \text{ is} \\ \text{TELIC} : \text{function of } x \\ \text{AGENTIVE} : \text{how } x \text{ came into being}\end{bmatrix}\end{bmatrix}$$

For the work presented here, TELIC role and AGENTIVE role are of the most interest. For example, for the noun *book*, *read* is a telic role verb and *write* is an agentive role verb.

Based on qualia structure, Pustejovsky [8] assume that the domain of individuals can be structured into three, increasingly complex, conceptual types:

a. Natural types: Natural kind concepts making reference only to Formal and Constitutive qualia roles; e.g. lion, rock, water, etc.

b. Artifactual types: Concepts making reference to Telic (purpose or function), or Agentive (origin). e.g. knife, beer, teacher, etc.

c. Complex types: Concepts making reference to a relation between at least two types from the other levels.

Pustejovsky [8] argues that a distinction between these two categories can be drawn by employing a notion of intentionality. In other words, it should be reflected in the telic and agentive roles. If no intentionality is involved, such words are natural types. On the contrary, artifacts are identified by the telic and agentive roles.

3.**Qualia relations in NCCVs.** NCCVs can be basically divided into three groups according to the part of the speech of the constituents in the compounds: VN, NV and VV. V stands for a verb or a verbal morpheme and N stands for a noun or a nominal morpheme. In this section, we will illustrate what qualia information (telic quale or agentive quale) the Vs can indicate and how in these three types of NCCVs.

3.1. **[VN] nominal compounds.** We will start with [VN] Nominal compounds, which are composed of two constituents. The first is a verb or verbal morpheme and the second is a noun or nominal morpheme. This construction is modifier-headed with the head on the right. There are four qualia relations to be found in this nominal compound construction.

3.1.1. **V is the TELIC of the compound.** In the compound *zheng-guo*, the modifying verb *zheng*'steam' relates to *guo*'s purpose, which is to steam food. The preferred interpretation of this compound is that it is a pot for steaming food. The verb *zheng* is the telic role of the nominal compounds as shown in (2).

$$(2)\begin{bmatrix}\text{zheng - guo} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{tool}] \\ \text{QUALIA} = [\text{TELIC} = \text{zheng' steam'}]\end{bmatrix}$$

TABLE 1.V is the TELIC of the [VN] compound

zheng-guo	蒸锅	steam-pot	steamer/a pot for steaming food
chuan-piao	传票	summon-ticket	summons
kan-dao	砍刀	chop-knife	hacking knife/chopper
jiuhu-che	救护车	rescue-car	ambulance
wo-shi	卧室	lie-room	bedroom
yinyong-shui	饮用水	drink-water	drinking water

More examples are presented in table 1. All the head nouns are artifactual types except *shui*'water', which is a natural type. However, the compound noun *yinyong-shui* is an artifactual type because the verb *yinyong*'drink' gives a function (i.e., a Telic value) to the head noun *shui* and change it to an artifactual type [9]. Therefore, all the compounds of this type are artifactual types.

3.1.2. **V is the AGENTIVE of the compound.** The modifying verbs relate to the origin or bringing about of the objects described by the compounds in table 2. For example, the verb *zhi*'*weave'* and the verb *ku*'cry' indicate the agentive roles of the compound *zhi-jin* and *ku-sheng* respectively, because brocade is woven of silk and cry is produced by crying. l*ai-gao* and *lai-xin* are special because the verb *lai* 'come' is an unaccusative verb, which has nothing to do with intensionality. All the head nouns and compounds in table 2 are artifactual types except *sheng* and *ku-sheng*.

$$(3)\begin{bmatrix} \text{zhi - jin} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{phys\_obj}] \\ \text{QUALIA} = [\text{AGENTIVE} = \text{zhi' weave'}] \end{bmatrix}$$

TABLE 2. V is the AGENTIVE of the [VN] compound

zhi-jin	织锦	weave-brocade	brocade/ picture-weaving in silk
zuo-wen	作文	write-composition	composition
lai-xin	来信	come-letter	incoming letter
lai-gao	来稿	come-manuscript	incoming manuscript
hua-xiang	画像	draw-portrait	portrait
diao-xiang	雕像	carve-portrait	statue
chao-mian	炒面	fry-noodle	fried noodle
shu-ming	署名	sign-name	signature
ku-sheng	哭声	cry-sound	cry

3.1.3. **VN is the TELIC of the compound.** The nominal compounds in table 3 are all deverbal nouns from verbal compounds or verbal phrases whose structure is verb-object. Interestingly, the verbal compounds or verbal phrases indicate the purpose or function of the noun form. For example, the compound *guan-jia* denotes a person who manages a house and the literal meaning of the verb phrase *guan jia* is exactly *managing a house*. Similarly, *sha-che* refers to a tool which is used to stop a car (i.e. *sha che*). All the compounds of this type are artifactual types.

(4) $\begin{bmatrix} \text{sha - che} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{tool}] \\ \text{QUALIA} = [\text{AGENTIVE} = \text{sha che' stop a car'}] \end{bmatrix}$

TABLE 3. VN is the TELIC of the [VN] compound

guan-jia	管家	manage-house	housekeeper
sha-che	刹车	stop-car	brake
tao-xiu	套袖	encase-sleeve	sleevelet/sleeve-covering
bang-tui	绑腿	bind-leg	leg wrappings
dou-du	兜肚	carry-abdomen	an undergarment covering the chest and abdomen
dian-jian	垫肩	pad-shoulder	shoulder pad/shoulder padding
wei-qiang	围墙	enclose-wall	enclosure

3.1.4. **VN is the AGENTIVE of the compound.** The compounds in table 4 are also deverbal nouns from verbal compounds or verbal phrases whose structure is verb-object. However, unlike the nominal compounds described in section 3.1.3, the verbal compounds or verbal phrases do not show the telic roles but the agentive roles of the nominal compounds. For instance, a paper-cut is created by cutting paper. The verbal phrase *jian zhi* is the TELIC of the nominal compound *jian-zhi*. All the compounds of this type are artifactual types.

(5) $\begin{bmatrix} \text{jian - zhi} \\ \text{ARGSTR} = [\text{ARG}_1 = \text{phys\_obj}] \\ \text{QUALIA} = [\text{AGENTIVE} = \text{jian zhi' cut paper'}] \end{bmatrix}$

TABLE 4. VN is the AGENTIVE of the [VN] compound

jian-zhi	剪纸	cut-paper	paper-cut/scissor-cut
jian-bao	剪报	cut-newspaper	cutting/ newspaper clipping

3.2. **[NV] compounds.** Nominal compounds of this type are composed by two elements. The first is a noun or nominal morpheme and the second is a verb or a verbal morpheme. We identify three qualia relations in this type of compounds.

TABLE 5. NV is the AGENTIVE of the [NV] compound

tian-fu	天赋	god-give	natural gift
xin-de	心得	heart-get	what one has learned from work, study etc.
huo-shao	火烧	fire-burn	baked wheaten cake
she-tui	蛇蜕	snake-exuviate	exuviate
chan-tui	蝉蜕	cicada-slough	cicada slough

3.2.1. **NV is the AGENTIVE of the compound.** The compounds in table 5 are deverbal

nouns from verbal compounds or verbal phrases having a subject-predicate structure. The verbal compounds or verbal phrases show the agentive quale of the deverbal nouns. For example, the literal meaning of *tian-fu* is a gift given by the god. All the compounds of this type are natural types.

3.2.2. **V is the AGENTIVE of the nominal.** The compounds in table 6 are right-headed constructions and the verbs on the right functions as nouns. In nominal compounds like *shi-diao*, *diao* 'to carve' is a verb but it occupies the right-hand position of the noun. It has been reanalyzed and functions as a noun although it cannot be used as a noun independently [6]. *Shi* modifys the noun form of *diao* and the verb form of *diao*'to carve' reveals that *shi-diao* is created by carving stones. While *diao* is the agentive role of the compound, *shi* indicates the constitutive quale by specifying the material of the compound. Besides *diao*, *ke* is another productive verb receiving a noun form class interpretation by virtue of its being the right-hand member of a gestalt noun. All the compounds of this type are artifactual types.

TABLE 6. V is the AGENTIVE of the [NV] compound

shi-diao	石雕	stone-carve	carved stone
bei-diao	贝雕	shell-carve	shell carving
qi-diao	漆雕	lacquer-carve	carved lacquerware
ya-diao	牙雕	ivory-carve	ivory carving
bing-diao	冰雕	ice-carve	ice sculpture /ice carving/ carved ice
shi-ke	石刻	stone-care	carved stone
mu-ke	木刻	sood-carve	woodcut

TABLE 7. V is the TELIC of the [NV] compound

ya-shua	牙刷	tooth-brush	toothbrush
fa-shua	发刷	hair-brush	hair brush
guo-gai	锅盖	pot-cover	pot cover
ping-gai	瓶盖	bottle-cover	cap
xiu-tao	袖套	sleeve-encase	sleevelet/sleeve-covering
bei-tao	被套	quilt- encase	bed sack
bei-zhao	被罩	quilt-cover	quilt cover
mian-zhao	面罩	face-cover	face guard
hua-zhan	画展	art-exhibit	art exhibition
shou-kao	手铐	hand-lock	handcuffs
xie-dian	鞋垫	shoe-pad	shoe-pad
wendu-ji	温度计	temperature-measure	themometer
xueya-ji	血压计	blood pressure-measure	tonometer
bi-xi	笔洗	writing-brush-wash	writing-brush washer small tray for washing brushes

3.2.3. **VN is the TELIC of the nominal.** Similar to the compounds in table 6, the verbs in the compounds in table 7 also undergo reanalysis because they occupy the right-hand position of nouns. But, unlike *ke*, some of these verbs can be used as a noun by adding a suffix *-zi (-子)* to them, e.g. *dianzi, shuazi, gaizi* and *zhaozi.* These compounds are also right-headed constructions. Interestingly, the reverse forms of the compounds are verb-object verbs or verbal phrases and can specify the purposes of the VN nominal compounds. For instance, *ya-shua* is something to brush teeth (i.e. *shua ya*) and *bi-xi* is used to wash writing-brush (i.e. *xi bi*). It is interesting that t*ao-xiu* in table 3 and *xiu-tao* in table 7 has the same meaning. The only difference between them is the order of *tao* and *xiu.* The former is VN compound and the latter is NV compound. There are some productive patterns such as N-*shua*, N-*dian*, N-*zhao* N-*ji* and N-*zhan*. All the compounds of this type are artifactual types and the vast majority of these compounds are tool nouns.

3.3. **[V_1V_2] nominal compounds.** The nominal compounds of this type are from verbal compounds or phrases, which are composed of two verbs or verbal morphemes and have a structure of coordination or verb-object.

3.3.1. **V_1V_2 is the TELIC of the nominal compound.** It can be seen from table 8 that the verbal form of V_1V_2 reveals the function or purpose of the nominal compound. For instance, *bian-ji* denotes persons who specially compile (*bian*) and edit (*ji*) books, magazines or something like this. *kai-guan* is a tool which is used to open(*kai*) and shut(*guan*) something like radios. *dao-yan* is a person who directs(*dao*) one's act(*yan*). Both *bian-ji* and *kai-guan* have a structure of coordination. The structure of *dao-yan* is verb-object. All the compounds of this type are artifactual types.

TABLE 8. V_1V_2 is the TELIC of the [V_1V_2] compound

bian-ji	编辑	compile-edit	editor
cai-gou	采购	select-purchase	a person who makes purchases for an organization or enterprise
kan-hu	看护	look-protect	nurse/look after
cai-feng	裁缝	cut-sew	tailor
zhuang-shi	装饰	adorn-decorate	adornment
pu-gai	铺盖	spread-cover	bedding/bedclothes
kai-guan	开关	open-shut	switch
bu-tie	补贴	subsidize-allowance	subsidy
dao-yan	导演	dierct-act	director
ban-wu	伴舞	accompany-dance	dancing partner

TABLE 9. V_1V_2 is the AGENTIVE of the [V_1V_2] compound

diao-ke	雕刻	carve-engrave	grave
diao-su	雕塑	carve-mould	sculpture
chuang-zuo	创作	create-write	creation
jian-zhu	建筑	build-construct	building
gong-xian	贡献	contribute-dedicate	contribution
shang-ci	赏赐	reward-grant	award
ji-lu	记录	write-record	record

3.3.2. **V_1V_2 is the AGENTIVE of the nominal.** It can be seen from table 9 that the verbal form of V_1V_2 reveals the origin or the bringing about of the nominal compound. For example, *diao-ke* is created by carving (*diao*) and engraving (*ke*). All the compounds of this type are artifactual types.

3.4. **Summary.** To sum up, generally speaking, there are three constructions to form NCCVs: VN, NV and V_1V_2. The verbal elements in these constructions can indicate the telic roles or agentive roles of the compounds in different ways as illustrated in table 10. Nearly all the NCCVs are artifactual types and most of the head nouns are artifacutal types too.

TABLE 10. Qualia relations in NCCVs

VN	V is TELIC	wo-shi 卧室 zheng-guo 蒸锅 yinyong-shui 饮用水	modifier-head
	V is AGENTIVE	zhi-jin 织锦 ku-sheng 哭声	modifier-head
	VN is TELIC	guan-jia 管家 sha-che 刹车 tao-xiu 套袖	deverbal nouns
	VN is AGENTIVE	jian-zhi 剪纸 jian-bao 剪报	deverbal nouns
NV	NV is AGENTIVE	tian-fu 天赋 chan-tui 蝉蜕	deverbal nouns
	V is AGENTIVE	shi-diao 石雕 mu-ke 木刻	modifier-head; V-to-N reanalysis
	VN is TELIC	ya-shua 牙刷 bi-xi 笔洗	modifier-head; V-to-N reanalysis
V_1V_2	V_1V_2 is TELIC	bian-ji 编辑 kai-guan 开关 dao-yan 导演	deverbal nouns
	V_1V_2 is AGENTIVE	diao-ke 雕刻 ji-lu 记录	deverbal nouns

By and large, VN construction is most productive and NV construction is next to it followed by V_1V_2 construction. Actually, deverbal nominal compounds are predominantly conventionalized and their productivity is much more restricted. Thus, the most productive construction is VN in which the V indicates the telic or agentive role. Many productive compounding patterns can be identified. Next section will examine these patterns.

4. **Productive compounding patterns.** VN nominal compounds are morphologically productive. A monosyllabic noun or nominal morpheme can be modified by a lot of verbs or verbal morphemes (especially disyllabic verbs) and form a group of nominal compounds. While some nominal morphemes are free and can occur as independent words, such as *ren(人)*, *che(车)* and *chang(厂)*, other nominal morphemes are bound, such as *shi(室)*, *jia(家)*, *ji(机)* and *wu(物)*. Some of the bound nominal morphemes are so productive that they look like suffixes. However, different from typical suffixes like *-zi(-子)* , *-er(-儿)* and *-tou(-头)*, these bound morphemes have clear semantic content. Above all, the content words which they are derived from can be retrieved. For example, *jia* is derived from the noun word *zhuan-jia(专家)* 'speacialist', *ji(*机) from *ji-qi(*机器*)* 'machine' and *wu(*物)from *wu-ti(*物体) 'object'. What's more, native Chinese speakers usually regard the compounds

formed by these nominal morphemes as right-headed nominal compounds rather than derived nouns. They are not typical suffixes and at best quasi-suffixes.

The most productive nominal morphemes can be divided into two groups. One is related with persons and the other is related with other things.

4.1. **Ns referring to person.** These nominal morphemes can be added to verbs or verbal morphemes to form agentive nominals. Pustejovsky [7] and Busa [2] identify two types of agentive nominals: individual-level nominals (ILNs) and stage-level nominals (SLNs). There are a series of distinctions between these two types of nominals. ILNs are associated with the telic role, expressing ability or capacity or a habit. For example, whereas a *smoker* refers to an individual who has the habit or the inclination to smoke with some of regularity, a *violinist* can be minimally characterized as an individual who has the ability of playing the violin. *Smoke* and *play violin* are telic roles of *smoker* and *violinist* respectively. On the other hand, SLNs express a temporal property and are associated with the agentive role, denoting the unique event of bringing about and proceeding the existence of an individual. For instance, it is the event of *traveling on vehicle* that makes a person a passenger. *Travel on vehicle* is the agentive role of *passenger*.

All the productive nominal morphemes of this type are equivalent to the English suffixes such as *-er,-or* or *-ist*. Unlike English suffixes, however, some of these nominal morphemes such as *yuan(员), jia(家), shi(师)*，*shou(手), shi(士)*，*sheng(生)* and *zu(族)* tend to form ILNs, but others such as *ren(人)* and *ke(客)* tend to form SLNs. In the ILNs, the verbs or verbal morpheme indicate the telic of the compounds. For example, *jiashi*'drive' shows the function of *jiashi-yuan*'driver' and *du*'read' indicates the bringing about of *du-zhe* 'reader'. More examples are presented in table 11 and 12. Note that *zhuchi-ren* 'compere' is an ILN.

TABLE 11. V is the TELIC of the ILN

jiashi-yuan	驾驶员	drive-personnel	driver
bang-shou	帮手	help-hand	helper
sheji-shi	设计师	design-master	designer
shoucang-jia	收藏家	collect-specialist	collector
chuanjiao-shi	传教士	preach-bachelor	preacher
fuwu-sheng	服务生	serve-student	server
shangban-zu	上班族	work-race	office worker
zhuchi-ren	主持人	host-person	host

TABLE 12. V is the AGENTIVE of the SLN

daiyan-ren	代言人	Speak for-person	spokesperson
faxing-ren	发行人	publish-person	publisher
cheng-ke	乘客	ride-guest	passenger
you-ke	游客	visit-guest	visitor

4.2. **Ns referring to other things.** The statistics of the distribution of the two quales (i.e. TELIC and AGENTIVE) within compounds shows that they are used with apparently different frequencies. It is clear from table 13 that the productive patterns prefer TELIC to

AGENTIVE: Most nominal morphemes, namely *qi(器)*, *shi(室)*, *chang(场), zhan(站), chang(厂), yuan(院), zhi(纸), yao(药)* and *yi(仪)* only select TELIC Vs. Out of 75 VN compounds containing *ji(机)* in Hownet dictionary, 93% of these are TELICselecting. All the Vs in the compounds containing *sheng* are agentive roles of the compounds. *pin(品)*and *wu(物)* can be modified by telic Vs as well as agentive Vs. Specific examples are given in table 14. A careful examination reveals that the productive morphemes preferring TELIC Vs are artifactual types, e.g. *chang(场)* and *ji(机),* while those preferring AGENTIVE Vs are natural types, e.g. *sheng(声)* and *wu(物)*. In other words, nouns or nominal morphemes of artifactual types are more inclined to select TELIC Vs than those of natural types.

TABLE 13. The distribution of the TELIC and AGENTIVE in VN compounding patterns

Patterns	Total[1]	TELIC	AGENTIVE	Other
V-qi(器) 'implement'	63	63=100%	0	0
V-shi(室) 'room'	18	18=100%	0	0
V-dao(刀) 'knife'	17	17=100%	0	0
V-chang(场) 'field'	16	16=100%	0	0
V-zhan(站) 'station'	14	14=100%	0	0
V-chang(厂) 'factory'	13	13=100%	0	0
V-yuan(院) 'yard'	8	8=100%	0	0
V-zhi(纸) 'paper'	6	6=100%	0	0
V-yao(药) 'medicine'	6	6=100%	0	0
V-yi(仪) 'instrument'	4	4=100%	0	0
V-ji(剂) 'agent'	33	30=90%	3=10%	0
V-ji(机) 'machine'	75	70=93%	0	5=7%
V-che(车) 'vehicle'	22	16=73%	5=23%	1=4%
V-chu(处) 'place'	7	5=71%	0	2=29%
V-pin(品) 'article'	35	21=60%	12=34%	2=6%
V-wu(物) 'thing'	19	8=42%	8=42%	3=16%
V-sheng(声) 'sound'	9	0	9=100%	0

TABLE 14. the quale in VN productive patterns

chucang-shi	储藏室	store-room	storeroom	TELIC
liangang-chang	炼钢厂	make steel-factory	steel plant	TELIC
dayin-ji	打印机	print-machine	Printer	TELIC
tiaowei-pin	调味品	flavor-article	flavoring	TELIC
fangzhi-pin	纺织品	weave-article	textile	AGENTIVE
shi-wu	食物	eat-thing	food	TELIC
chuban-wu	出版物	publish-thing	publication	AGENTIVE
jiao-sheng	叫声	cry-sound	cry	AGENTIVE

[1] Only disyllabic and trisyllabic compounds are examined.

The fact that TELIC patterns are more productive can be explained. The telic role encodes the characterizing function or purpose of an entity, while the agentive role expresses the coming into being of an entity. Clearly, real artifactual concepts such as "knife", "room" and "machine" are intentionally defined by reference to both AGENTIVE and TELIC [8]. In some sense, TELIC is more important than AGENTIVE because every artifactual object is created for some specific purpose. In fact, for such objects, it is difficult to imagine creation without purpose. The telic quales are usually not shared by other members of the same category. For instance, various knives have a quite different purpose (e.g. *to chop, razor, scissor, stab, scrape*) and therefore the category of "knife" can be subcategorized by these specific purposes, as seen in Chinese VN compounds like *kan-dao*, *ti-dao*, *jian-dao*,*ci-dao* and *gua-dao* (see table 15 and figure 1). The verbs modify the specific purpose of the noun *dao* and distinguish it from other sorts of knives. For example, in *kan-dao*, the verb *kan*'chop' distinguishes *dao* from other sorts such as *ti-dao*. *kan-dao* refers to a subtype of *dao* 'knife' used to chop something. Likewise, *ti-dao* refers to a subtype of *dao* used to raze (*ti*) something. In a word, telic quale plays more significant role in subcategorization than agentive quale.

TABLE 15. V-*dao*

kan-dao	砍刀	chop-knife	chopper
ti-dao	剃刀	raze-knife	razor
jian-dao	剪刀	scissor-kinfe	scissors/ shears
gua-dao	刮刀	scrape-knife	scraper/ scraping cutter
ci-dao	刺刀	stab-knife	bayonet

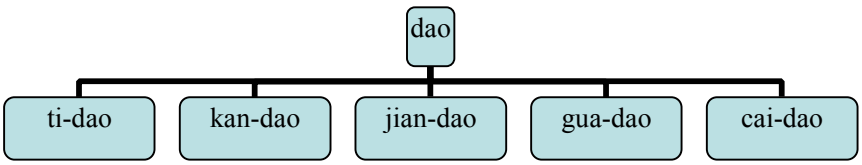

FIGURE 1. The subcategorization of *dao*

5. Differences between Chinese and English in word formation. Numerous examples described above show many differences between Chinese and English in word formation. First of all, Chinese uses compounding process to create new words more often than English. It can be seen from the fact that many compounds in Chinese cannot be translated as compounds in English, but derived nominals or deverbal nouns. For example, the English equivalent of *kan-dao* (砍刀 literally, chop-knife) is *chopper*, a derived nominal. *Ku-sheng*(哭声 literally, cry-sound) is translated as *cry*, a deverbal noun.

Secondly, VN nominal compounds in Chinese are often translated as NN compounds in English for two reasons. First, in English a verb cannot modify a noun directly and it must become a gerundive nominal before modifying a noun. For instance, Chinese VN compound *yinyong-shui* (饮用水 literally, drink-water) is translated as *drinking water* rather than *drink water*. Second, compared to English, Chinese compounds are in favor of VN construction rather than NN construction as in words like *zao-zhi-chang* (造纸厂 literally,

make-paper-mill) ‘paper mill’ and *wo-shi* (卧室 literally, lie-room) ‘bed room’. Although *zhi-chang* is good in Chinese, it is less common than *zao-zhi-chang*. In Peking University modern Chinese corpus, the number of occurrences of *zhi-chang* is 136 and that of *zao-zhi-chang* is 672.

Thirdly, as shown in (6), a tri-syllabic Chinese VN compound may have a different internal structure from its English equivalent, an NN compound.

(6) Chinese VN compound: (noun$_1$+verb)+noun$_2$
English NN compound: noun$_2$+(verb+*er)/* noun$_2$+verbal noun

Let’s take *ge-cao-ji* (割草机 literally, cut-hay-machine) for example. Its English equivalent is *hay cutter*. They correspond to different structures as illustrated in figure 2 and figure 3.

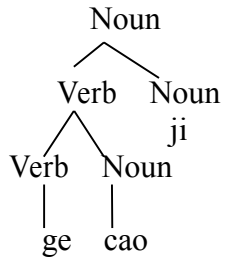

FIGURE 2. The internal structure of *ge-cao-ji*

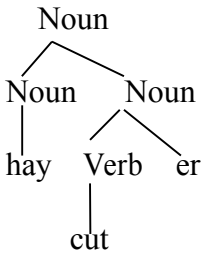

FIGURE 3. The internal structure of *hay cutter*

In the first structure, the verb *ge* combines with the noun *cao* to form an verbal compound *ge-cao*. Then the verbal compound combines with the noun *ji* to form the nominal compound *ge-cao-ji*. In the second structure, however, the verb *cut* combines with the suffix *-er* resulting in a noun *cutter*. Then the noun *cutter* combines with the noun *hay* to form the nominal compound *hay cutter*. The (verb+er) position may be a deverbal noun as in words like *bicycle park*, whose Chinese equivalent is *cun-che-chu* (存车处 literally, store-bicycle-place)

These differences between Chinese and English in word formation are due to a general difference between isolating languages and inflected languages. Isolating languages want agglutinative inflectional markers, while inflected languages have a lot of derivational affixes. Compounding is regarded as one means of word formation especially productive for isolating languages such as Chinese.

6. **Application.** The analysis of qualia information in NCCVs presented in this paper could

be potentially useful for a variety of NLP tasks. One obvious application is natural language understanding. Pustejovsky [7] looks at the relative transparency or opacity of how a noun refers as expressed in the value of the noun's qualia structure. He refers to this property as referential transparency and defines it as follows:

(7) The REFERENTIAL TRANSPARENCY of an NP is a characterization of the specificity of qualia values for that NP.

(i) If an NP is only weakly constrained by the values of its qualia, then it is referentially opaque.

(ii) If an NP is strongly constrained by the values of its qualia, then it is referentially transparent .

Referential transparency can captures the intuitive distinction between the informativeness of the nominal types. The more transparent a noun is, the more information it can provide. There is a range of transparency and these are not absolute categories, except in the case of pronouns such as *she* and proper nouns such as *Mary*, which are referentially opaque NPs, in that there is no mode of explanation (i.e. quale) suggesting how this term denotes. The other end of the spectrum are nouns such as *baker*, *lecturer* and *bread*. We argue that NCCVs such as *zuo-jia*(literally, write-specialist)'writer' and *zuo-pin* (literally, write-article) 'writing' are more referential transparent than other nouns such as *shu*'books', in that the telic role of *zuo-jia* and the agentive role of *zuo-pin* are indicated literally by the modifying verb (i.e., *zuo*'write'). The degree of referential transparency of a noun will determine the default interpretation of how a phrase is semantically reconstructed within an ambiguous environment. Consider the following examples.

(8) a. 张三 的 书
ZhangSan de shu
ZhangSan POSS book
'ZhangSan's books'

b. 这 位 作家 的 书
Zhe wei zuojia de shu
the CL writer POSS book
'The writer's books'

c. 这 位 作家 的 作品
Zhe wei zuojia de zuopin.
the CL writer POSS writing
'The writer's works'

The default interpretation of (1a) is *the books belonging to ZhangSan* because the [NP_1 *de* NP_2] construction usually expresses a relation of possession. Although the possession interpretation is also available for (1b), its default interpretation is *the books written by the writer*, because the telic role from the agentive nominals seems to override the constructional meaning, namely possession meaning. (1c) has only one interpretation: *the works written by the writer*. The possession meaning has been ruled out because both the subject and the complement are NCCVs and the verb *zuo* in these two compounds override

the construction meaning completely. (9a) is an example of complement coercion. *Gan* 'rush' selects event function as its complement, but *shu*'book' cannot satisfy its requirement. Therefore, *gan* coerces *shu* to assume an event type. The agentive role of the complement noun *shu* makes it possible. The default interpretation of (9a) is *to rush to finish writing the book*. However, the reconstructed event in (9b) is not *writing the paper* but *editing the paper*. It is because the *bian-ji* (literally, compile-edit) is an NCCV and more transparent than *shu* in that its literal meaning can reveal its telic role (i.e. edit) (see 3.3.1) .

(9) a.我　正在　赶　这　本　书。
wo zhengzai gan zhe ben shu.
I being rush the CL book
'I am rush to finished *writing the book*.'
b.编辑　正在　赶　这　本　书。
bianji zhengzai gan zhe ben book.
editor being rush the CL book
'The editor is rushing to finish *editing the book*. '

In a word, the high degree of transparency of NCCVs can help disambiguation. Additionally, it can reduce the degree of difficulty of text comprehension. The more NCCVs a text contains, the easier it is to understand since NCCVs are much more transparent than other nominal compounds. For example, it is much easier to interpret the VN compounds than NN compounds in that there is a covert predicate to be reconstructed in the paraphrase of the latter while the predicate is overt in the former. E.g., *zaozhi-chang* (造纸厂) is much easier to understand than *zhi-chang*（纸厂）.

Another important use is in the process of new words or Unknown Words (UWs). UWs are always a challenge for NLP. As well as known, some compounding patterns are very productive and it is impossible and uneconomic to add all the compounds to the lexicon. The optimal arrangement will be to list frequent and idiosyncratic compound forms in the lexicon and use the compositional rules for forms which are not listed in the dictionary since their meaning is compositional. A given compounding pattern usually has a paraphrase pattern[2]. For example, V-*chang*（厂） means *a factory which is used for V* since the V always denotes the telic role of the compound. E.g., *zaozhi-chang* means *a factory which is used for making paper* (*zao zhi*). The paraphrase pattern can help in the automatic interpretation of UWs. *Niang-jiu-chang* (酿 酒 厂 literally, make-alcoholic beverages-factory) is not listed in the Hownet Dictionary, but its meaning can be predicted based on the paraphrase pattern, namely, *a factory which is used for making alcoholic beverages.*

Besides, this work also has important consequences for applications in multilingual natural language processing, especially English-Chinese translation. It is important to note, however, that not all Chinese NCCVs can be translated as NCCVs in English as mentioned in section 6. Neither translation from English to Chinese nor the reverse is straightforward. In this sense, it is better for the Chinese-English dictionary for machine translation to

[2] See [10] for further discussion about productive compounding patterns and their paraphrase patterns.

include as many NCCVs as possible.

7**. Conclusion.** Under the theoretical framework of qualia structure, this article addresses the semantic relations of NCCVs in Chinese. It turns out that the verbal elements in different constructions can indicate the telic role or agentive role of the nominal compounds in different ways. Nearly all the NCCVs are artifactual types and most of the head nouns are artifactual types too. The VN compound construction in which the V indicates the telic is highly productive and many productive compounding patterns can be identified. This study also reveals many differences between Chinese and English in word formation. The analysis of qualia information in NCCVs presented in this paper could be potentially useful for a variety of NLP tasks such as natural language understanding (especially disambiguation), automatic interpretation of unknown words and Chinese-English machine translation.

Acknowledgment. This work is partially supported by the National Social Science Foundation of China (No. 10CYY032), Fundamental Research Funds for the Central Universities and National Natural Science Foundation of China (No. 61103089).

REFERENCES

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

[2] Busa, Federica. The semantics of agentive nominals in the Generative Lexicon. In Patrick Saint-Dizier (ed.), *Predicative Forms in Natural Language*. Amsterdam: Kluwer. 1997.

[3] Johnston, M. and Busa, F. Qualia structure and the compositional interpretation of compounds. In E. Viegas (ed.), *Breadth and Depth of Semantics Lexicons,* 167-187. Dordrecht: Kluwer. 1999.

[4] Lee, Chih-yao, Chia-hao Chang, Wei-chieh Hsu & Shu-kai Hsieh. Qualia modification in noun-noun compounds: A Cross-Language Survey. *Proceedings of the 22nd Conference on Computational Linguistics and Speech Processing (ROCLING-2010),* 379-390. 2010.

[5] *Li*, Charles N. *& Thompson*, S. A. *Mandarin Chinese: A Functional Reference Grammar.* Berkeley: University of California Press.1981.

[6] Packard, J.L. *The Morphology of Chinese*. New York: Cambridge University Press. 2000.

[7] Pustejovsky, James. *The Generative Lexicon*. Cambridge:The MIT Press. 1995.

[8] Pustejovsky, James. Type construction and the logic of concepts. In Federica Busa and Pierrette Bouillon (eds.), *The Syntax of Word Meanings*, 91-123. Cambridge: Cambridge University Press. 2001.

[9] Pustejovsky, James. Coercion in a generative theory of argument selection. *Linguistics* 49:1401-1431. 2011.

[10] Qiu, Likun(邱立坤). Interpretation mode of mono-syllable nouns and semantic structure relation of tri-syllable nouns. *Proceedings of the 9th National Conference of China on Computational Linguistics*, 203-208. 2007.

[11] Song, Zuoyan(宋作艳). On the semi-affixes which can trigger event coercion. *Chinese Teaching in the World* 4: 446-458. 2010.