CSCanada
**Canadian Social Science**
Vol. 14, No. 3, 2018, pp. 19-28
**DOI:**10.3968/10193

ISSN 1712-8056[Print]
ISSN 1923-6697[Online]
www.cscanada.net
www.cscanada.org

# An Empirical Study of English-Chinese Translation of Novel Context-Free Compound Nouns and Phrases

WANG Wei[a],*; ZHOU Weihong[b]

[a]School of Interpreting and Translation, Beijing International Studies University, Beijing, China.
[b]School of Foreign Languages, Ludong University, Yantai, China.
*Corresponding author.

Received 13 December 2017; accepted 4 February 2018
Published online 26 March 2018

## Abstract

The current study designs a compound translation test and finds out that unlike English speakers, Chinese translators tend to bypass syntactic paraphrase and directly conduct semantic processing on the surface structure of compounds/phrases. Syntactic operations, semantic categories, and world knowledge are important factors in compound interpretation and translation. Syntactic analysis and semantic processing are important factors in the process of interpretation and translation of novel context-free compounds and phrases. The present study also reveals the psychological differences between English and Chinese speakers. Syntactic transformation knowledge is also quite helpful in disambiguating compounds/phrases with the same surface structures. Statistical results demonstrate that the abstractness of compounds affect translators' processing effort as well as accuracy. Other possible factors in compounding comprehension and translation include world knowledge, contextual information and pragmatic awareness.

**Key words:** Translation; Context-free; Compound nouns; Phrases; Synthetic

Wang, W., & Zhou, W. H. (2018). An Empirical Study of English-Chinese Translation of Novel Context-Free Compound Nouns and Phrases. *Canadian Social Science, 14*(3), 19-28. Available from: http://www.cscanada.net/index.php/css/article/view/10193
DOI: http://dx.doi.org/10.3968/10193

## INTRODUCTION

### A. Why Choosing Compound Nouns and Phrases?

A compound can be defined as "the formation of a new lexeme by adjoining two or more lexemes" (Bauer 2003, p.40). Compound nouns derived from the deep structure of VERB + OBJECT (cf. Bloomfield, 1933, 231-232; Marchand, 1969, pp.15-19) are usually labeled as synthetic compound nouns or deverbal compounds. While Roeper & Siegel (1978) adopt the notion of "verbal compounds" i.e., "those with a verbal affix: -er, -ing, -ed" and discussed lexical transformation for verbal compounds. Lexical transformation can generate synthetic compound nouns from verbal phrases, e.g., the verbal phrase COOK RICE can generate "rice cooker" (compound noun) and "rice cooking" (noun phrase). Similarly, the verbal phrase LAND ON COMET can generate "comet landing" (noun phrase). TAKE IN SALT also generates the phrase "salt intake". However, up to the present days linguists still have not found a clear-cut boundary between compounds and phrases. So it is hard to distinguish compounds from phrases with same syntactic structures. Accordingly, it is quite difficult for average English speakers to tell the reasons why "rice cooker" can be regarded as a compound but "salt intake" is a phrase. Lieber & Štekauer (2009) summarize various criteria of defining compounds and make the conclusion "that there are (almost) no reliable criteria for distinguishing compounds from phrases or from other sorts of derived words." Chen Ping (2016 academic communication) believes that "whether a word/phrase is compiled by mainstream dictionaries can be used as a criterion to identify a compound." We may find that there is a rather vague distinction between synthetic compounds and deverbal phrases. At one end of the continuum, we may find typical compounds (e.g. *fish* *farming, bus driver*); at the opposite end, every day newly coined deverbal phrases (e.g. *bike share, bike sharing, shared-bike hunter, comet landing, drone crackdown*) flash in.

### B. Significance of the Issue

This principle of word formation is marked with the feature of uniformity and productivity, i.e., it generates compounds with suffixes: *-er, -or, -ing* and zero suffixes (e.g. *shot block, drone crackdown, bike share, salt intake*). Therefore, synthetic compound nouns are quite productive. Hence the present study proposes to investigate the cognition and translation of both synthetic compounds as well as other deverbal phrases, which have great significance in compound learning and translator/interpreter training. Compound interpretation has practical applications such as machine translation (Cao & Li, 2002; Baldwin & Tanaka, 2004) and AI applications such as "automatic recognition of relations between pairs of nominals in a sentence" (Girju et al., 2009). However, its theoretical importance and practical potentials have long been ignored for certain reasons.

### C. Literature Review

Lauer and Dras (1994) point out that there are three components to compound cognition: identification of the compound from within the text, syntactic analysis of the compound (left versus right association), and the interpretation of the underlying semantics. Accordingly the present study follows the model of Lauer and Dras (1994) that there are three major phases in compound cognition for English speakers, including: Phase 1, identification; Phase 2, syntactic analysis; Phase 3, semantic interpretation. As for average English speakers, they may transform synthetic compound nouns into the corresponding verbal phrases without much processing effort, featuring syntactic processing is automatic and deeply rooted in their I-language.

Roughly speaking, there are two perspectives in analyzing compound compositionality: syntactic analysis (Adams, 1973; Liberman & Sproat, 1992; Lauer, 1995) and semantic analysis (Downing, 1977; Warren, 1978; Levi, 1978). Levi's (1978) scale of compound semantic relations is a little bit small. While some linguists (Jespersen, 1954; Downing, 1977) believe that the number of compound semantic relations could be infinite. In Warren's (1978) opinion, there are six major hierarchical semantic relations. Nastase and Szpakowicz (2003) divide five coarse-grained super-relations (CAUSALITY, QUALITY, TEMPORALITY, PARTICIPANT, SPATIAL) and further elaborate thirty fine-grained relations (CAUSE, EFFECT, PURPOSE, DETRACTION; CONTAINER, CONTENT, EQUATIVE, MATERIAL, MEASURE, TOPIC, TYPE; FREQUENCY, TIME AT, TIME THROUGH; AGENT, BENEFICIARY, INSTRUMENT, OBJECT, OBJECT PROPERTY, PART, POSSESSOR, PROPERTY, PRODUCT, SOURCE, STATIVE, WHOLE; DIRECTION, LOCATION, LOCATION AT, LOCATION FROM). Compared with other models, the semantic classification proposed by Nastase and Szpakowicz (2003) is more exhaustive and inclusive. Thus the present study adopts this classification of semantic relations and manually annotated the compounds and phrases semantically. For English speakers, compound "attachments are not syntactically, but semantically governed" (Girju et al., 2005). Štekauer (2009) holds a similar viewpoint – "most predictable readings are motivated by the semantic components representing prototypical features of the motivating objects." Xu (2014) conducts an experiment to investigate the role of structural categories in evaluating semantic transparency (Sandra, 1990; Zwisterlood, 1994; Libben, 1998) of Chinese compounds and deems it as an influential factor in meaning prediction and interpretation. But few linguists and scholars have ever touched the issue of Egnlish compounds interpretation and translation in China mainland.

### D. The Current Study

The current study attempts to investigate the relationship between old synthetic compounds and the newly emerged context-free deverbal compounds and phrases. Because of the syntactic similarity (OV syntactic pattern), we put forward Hypothesis 1 that there is significant correlation between the prediction and translation of old synthetic compounds and novel context-free synthetic compounds and phrases. We also propose Hypothesis 2 that unlike average English speakers, Chinese translators often neglect syntactic operations in their mental lexicon and come directly to Phase 3, i.e., semantic interpretation in the surface structure. Sometimes newly emerged compounds and phrases sharing the same surface structure (e.g., *shark warning* vs. *customer warning*) are actually syntactically different (O $N_V$ vs. N $N_V$), i.e., their deep structures are different. So we design 15 pairs of novel compounds and phrases in order to find out whether there are priming or constraining effects between each compound/phrase in the process of meaning prediction and translation.

## 1. METHOD

### 1.1 Training Data Collection

We extract instance compounds/phrases from two major sources. The 30 (15 compounds of OV-$_{er}$ and 15 compounds of OV-$_{ing}$) old synthetic compounds are randomly selected from *MacMillan English-Chinese Dictionary for Advanced Learners* (2005). As for the 30 novel compounds and phrases (15 compounds/phrases of "O $N_V$ , $N_A$ $N_V$" and 15 compounds/phrases of "N $N_V$"), they are chosen from the websites of BBC, CNN, FOX NEWS, NPR, BNC (British National Corpus), and COCA (Corpus of Contemporary American English). Each compound/phrase used in the test (see Appendix I) is manually tagged respectively according to syntactic and semantic categories.

### 1.2 Participants

Altogether 83 Chinese postgraduates majoring in translation and interpreting at École Supérieure d'Interprètes et de Traducteurs (School of Interpreting and Translation) of Beijing International Studies University participated in the test. All the subjects are 21 to 26 year-olds, and they all passed the national English proficiency test for English majors (TEM 8) in their undergraduate education years.

### 1.3 Sampling Procedures

The participants are required to translate the compounds and phrases into Chinese within 40 minutes. Their translated versions are marked by the five-point Likert scale ranging from 0 to 4.81 valid copies of testing paper are retrieved and the statistical data are processed with the software SPSS 20 (see Appendix II). Calculations include one-sample statistics of all testing items, Correlation between the prediction and translation of old synthetic compounds and novel context-free synthetic compounds/ phrases, mean values of each syntactic category, paired samples correlations, as well as paired samples of correlations.

## 2. RESULTS

In Table 1, the mean values for all instances are presented in the axis of Mean, ranging from the lowest (*draught excluder* .0864) to the highest (*problem-solving* 3.9877).

**Table 1**
**One-Sample Statistics**

| Syntactic categories | Compounds/ phrases | Mean |
|---|---|---|
| | *Cigarette lighter* | 3.7037 |
| | *Shipbuilder* | 2.3457 |
| | *Snowblower* | 1.6173 |
| | *Bread winner* | 2.7284 |
| | *Face-saver* | 1.0370 |
| | *Coffee maker* | 3.0247 |
| | *Dishwasher* | 3.5185 |
| OV-*er* | *Doorkeeper* | 3.0864 |
| | *Draught excluder* | .0864 |
| | *Earth mover* | 1.5802 |
| | *Fire-eater* | .7778 |
| | *Fire extinguisher* | 2.7284 |
| | *Fortune-teller* | 3.1111 |
| | *Glassblower* | .9877 |
| | *Goalkeeper* | 1.9753 |
| | *Asset-stripping* | 1.0123 |
| | *Coal mining* | 2.8519 |
| | *Decision-making* | 3.6914 |
| | *Housewarming* | .9506 |
| | *Housekeeping* | .9877 |
| | *Job-Sharing* | .4321 |
| | *Lawmaking* | 3.5556 |
| OV-*ing* | *Matchmaking* | 1.6914 |
| | *Blood poisoning* | 1.7284 |
| | *Peacekeeping* | 3.8272 |
| | *Problem-solving* | 3.9877 |
| | *Profit-sharing* | 2.3827 |
| | *Risk-taking* | 3.3333 |
| | *Trendsetting* | 2.2469 |
| | *Stamp collecting* | 3.9136 |

To be continued

 

Continued

| Syntactic categories | Compounds/ phrases | Mean |
|---|---|---|
| $N_A$ $N_V$ | *Comet landing* | .8519 |
| N $N_V$ | *Practice landing* | 1.2963 |
| O $N_V$ | *Shot block* | .5309 |
| N $N_V$ | *Stone block* | .8272 |
| O $N_V$ | *Drone crackdown* | .7037 |
| N $N_V$ | *Military crackdown* | 2.0247 |
| N $N_V$ | *Shark warning* | 1.8025 |
| O $N_V$ | *Customer warning* | 2.0123 |
| N $N_V$ | *Police shooting* | 2.8889 |
| $N_A$ $N_V$ | *Campus shooting* | 3.8395 |
| O $N_V$ | *Bike share* | 3.5802 |
| N $N_V$ | *Market share* | 3.1235 |
| O $N_V$ | *Australia attack* | 1.6420 |
| N $N_V$ | *Bomb attack* | 3.6296 |
| O $N_V$ | *Property Manager* | 1.7160 |
| N $N_V$ | *Assistant manager* | 2.6543 |
| O $N_V$ | *Trump accuser* | 1.9877 |
| N $N_V$ | *Nelly rape accuser* | 1.3333 |
| O $N_V$ | *Sleep deprivation* | 3.5679 |
| N $N_V$ | *Inner-city deprivation* | .5679 |
| O $N_V$ | *Salt intake* | 3.2222 |
| N $N_V$ | *Normal human intake* | 2.8148 |
| O $N_V$ | *Police shooting protests* | 2.0864 |
| N $N_V$ | *Labor protest* | 2.7531 |
| O $N_V$ | *Tax relief* | 3.5185 |
| N $N_V$ | *Food relief* | 2.7160 |
| O $N_V$ | *Seat reservation* | 3.4198 |
| N $N_V$ | *Computer reservation* | 2.2840 |
| O $N_V$ | *Star chase* | 3.4568 |
| N $N_V$ | *Police chase* | 2.6049 |

*Note.* Next sections will show the statistical results of the relationships between each variable tagged syntactically and semantically.

### 2.1 Correlation Between the Prediction and Translation of Old Synthetic Compounds and Novel Context-Free Synthetic Compounds/ Phrases

Table 2 provides the mean values for each syntactic category of compounds/phrases. Out of our expectation, the novel compounds/phrases with the structure of "O $N_V$ + $N_A$ $N_V$" (36.1358) and "N $N_V$" (33.3210) have higher mean values than the old compounds with the structure of "OV-*er*" under the same condition of de-contextualization. According to the metaphor cline (HUMAN > OBJECT > PROCESS > SPACE > TIME > QUALITY) proposed by Heine, Claudi and Hünnemeyer (1991, p.157), cognitive categories constitute a continuum (cline). Each category is more abstract than its neighbor on its right side. On the other direction, each one is more concrete than its neighbor on its right side. The mean value of OV-*ing* (36.5926) is higher than OV-*er* (32.3086), which is consistent with the metaphor cline. The category of PROCESS (OV-*ing*) requires less processing effort and has higher prediction rate than the categories of HUMAN and OBJECT (OV-*er*).

**Table 2**
**Mean Values of Each Syntactic Category**

| Syntactic categories | Mean values |
|---|---|
| OV-*er* | 32.3086 |
| OV-*ing* | 36.5926 |
| O $N_V$ + $N_A$ $N_V$ | 36.1358 |
| N $N_V$ | 33.3210 |

**Table 3**
**Paired Samples Correlations**

| | Syntactic categories | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | OV-er & O $N_V$ + $N_A$ $N_V$ | 81 | .405 | .000 |
| Pair 2 | OV-er & N $N_V$ | 81 | .506 | .000 |
| Pair 3 | OV-*ing* & O $N_V$ + $N_A$ $N_V$ | 81 | .443 | .000 |
| Pair 4 | OV-*ing* & N $N_V$ | 81 | .464 | .000 |
| Pair 5 | OV-er & OV-*ing* | 81 | .603 | .000 |
| Pair 6 | O $N_V$ + $N_A$ $N_V$ & N $N_V$ | 81 | .765 | .000 |

In Table 3, correlations of 6 pairs of syntactic categories are presented in the axis of Correlations with the Sig. (.000): "O $N_V$ + $N_A$ $N_V$ & N $N_V$" (.765) ＞ "OV-er & OV-*ing*" (.603) ＞ "OV-er & N $N_V$" (.506) ＞ "OV-*ing* & N $N_V$" (.464) ＞ "OV-*ing* & O $N_V$ + $N_A$ $N_V$" (.443) ＞ "OV-er & O $N_V$ + $N_A$ $N_V$" (.405). The statistical result confirms our first hypothesis, i.e., there is significant correlation between the prediction and translation of old synthetic compounds and novel context-free synthetic compounds and phrases. However, the correlation values cannot indicate the priming effect of syntactic categories because the correlation value of "O $N_V$ + $N_A$ $N_V$ & N $N_V$" (.765) is much higher than "OV-er & O $N_V$ + $N_A$ $N_V$" (.405), which means that the same syntactic structure does not increase the rate of predictability. In other words, the predictability rate is less syntactically motivated but more semantically motivated, confirming our second hypothesis that Chinese translators often bypass syntactic paraphrase and come directly to semantic processing.

### 2.2 Priming and Constraining Effects Between Compounds/Phrases With the Same Surface Structure

The current study tags the instances (OBJECT, DIRECTION, PROPERTY, MATERIAL, AGENT, TOPIC, LOCATION AT, INSTRUMENT) according to Nastase & Szpakowicz's (2003) semantic classification and statistical result is provided in Table 4 and Table 5.

**Table 4**
**Semantic Classifications**

| Syntactic categories | Semantic categories | Compounds/ phrases | Mean |
|---|---|---|---|
| | OBJECT | *Cigarette lighter* | 3.7037 |
| | OBJECT | *Shipbuilder* | 2.3457 |
| | OBJECT | *Snowblower* | 1.6173 |
| | OBJECT | *Bread Winner* | 2.7284 |
| | OBJECT | *Face-Saver* | 1.0370 |
| | OBJECT | *Coffee maker* | 3.0247 |
| | OBJECT | *Dishwasher* | 3.5185 |
| OV-*er* | OBJECT | *Doorkeeper* | 3.0864 |
| | OBJECT | *Draught excluder* | .0864 |
| | OBJECT | *Earth mover* | 1.5802 |
| | OBJECT | *Fire-eater* | .7778 |
| | OBJECT | *Fire extinguisher* | 2.7284 |
| | OBJECT | *Fortune-teller* | 3.1111 |
| | OBJECT | *Glassblower* | .9877 |
| | OBJECT | *Goalkeeper* | 1.9753 |
| | OBJECT | *Asset-stripping* | 1.0123 |
| | OBJECT | *Coal mining* | 2.8519 |
| | OBJECT | *Decision-making* | 3.6914 |
| | OBJECT | *Housewarming* | .9506 |
| | OBJECT | *Housekeeping* | .9877 |
| | OBJECT | *Job-sharing* | .4321 |
| | OBJECT | *Lawmaking* | 3.5556 |
| OV-*ing* | OBJECT | *Matchmaking* | 1.6914 |
| | OBJECT | *Blood poisoning* | 1.7284 |
| | OBJECT | *Peacekeeping* | 3.8272 |
| | OBJECT | *Problem-solving* | 3.9877 |
| | OBJECT | *Profit-sharing* | 2.3827 |
| | OBJECT | *Risk-taking* | 3.3333 |
| | OBJECT | *Trendsetting* | 2.2469 |
| | OBJECT | *Stamp collecting* | 3.9136 |
| $N_A$ $N_V$ | DIRECITON | *Comet landing* | .8519 |
| N $N_V$ | PROPERTY | *Practice landing* | 1.2963 |
| O $N_V$ | OBJECT | *Shot block* | .5309 |
| N $N_V$ | MATERIAL | *Stone block* | .8272 |
| O $N_V$ | OBJECT | *Drone crackdown* | .7037 |
| N $N_V$ | AGENT | *Military crackdown* | 2.0247 |
| N $N_V$ | TOPIC | *Shark Warning* | 1.8025 |

To be continued

 

Continued

| Syntactic categories | Semantic categories | Compounds/ phrases | Mean |
|---|---|---|---|
| O $N_V$ | OBJECT | *Customer warning* | 2.0123 |
| N $N_V$ | AGENT | *Police shooting* | 2.8889 |
| $N_A$ $N_V$ | LOCATION AT | *Campus shooting* | 3.8395 |
| O $N_V$ | OBJECT | *Bike share* | 3.5802 |
| N $N_V$ | PROPERTY | *Market share* | 3.1235 |
| O $N_V$ | OBJECT | *Australia attack* | 1.6420 |
| N $N_V$ | INSTRUMENT | *Bomb attack* | 3.6296 |
| O $N_V$ | OBJECT | *Property manager* | 1.7160 |
| N $N_V$ | PROPERTY | *Assistant manager* | 2.6543 |
| O $N_V$ | OBJECT | *Trump accuser* | 1.9877 |
| N $N_V$ | TOPIC | *Nelly rape accuser* | 1.3333 |
| O $N_V$ | OBJECT | *Sleep deprivation* | 3.5679 |
| N $N_V$ | LOCATION AT | *Inner-city deprivation* | .5679 |
| O $N_V$ | OBJECT | *Salt Intake* | 3.2222 |
| N $N_V$ | AGENT | *Normal human intake* | 2.8148 |
| O $N_V$ | OBJECT | *Police shooting protests* | 2.0864 |
| N $N_V$ | AGENT | *Labor protest* | 2.7531 |
| O $N_V$ | OBJECT | *Tax relief* | 3.5185 |
| N $N_V$ | PROPERTY | *Food relief* | 2.7160 |
| O $N_V$ | OBJECT | *Seat reservation* | 3.4198 |
| N $N_V$ | INSTRUMENT | *Computer reservation* | 2.2840 |
| O $N_V$ | OBJECT | *Star chase* | 3.4568 |
| N $N_V$ | AGENT | *police chase* | 2.6049 |

**Table 5**
**Means of Semantic Classifications**

| Semantic categories | Numbers | Mean |
|---|---|---|
| OLD OBJECT | 30 | 2.2967 |
| NEW OBJECT | 13 | 2.4188 |
| DIRECTION | 1 | .8519 |
| PROPERTY | 4 | 2.4475 |
| MATERIAL | 1 | .8272 |
| AGENT | 5 | 2.6173 |
| TOPIC | 2 | 1.5679 |
| LOCATION AT | 2 | 2.2037 |
| INSTRUMENT | 2 | 2.9568 |

**Table 6**
**Paired Samples Correlations**

| | Pairs | *N* | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | *Comet landing & practice landing* | 81 | .138 | .219 |
| Pair 2 | *Shot block & stone block* | 81 | .366 | .001 |
| Pair 3 | *Drone crackdown & military crackdown* | 81 | .337 | .002 |
| Pair 4 | *Shark warning & customer warning* | 81 | .306 | .006 |
| Pair 5 | *Police shooting & campus shooting* | 81 | .397 | .000 |
| Pair 6 | *Bike share & market share* | 81 | .048 | .668 |
| Pair 7 | *Australia attack & bomb attack* | 81 | .221 | .047 |
| Pair 8 | *Property manager & assistant manager* | 81 | .011 | .922 |
| Pair 9 | *Trump accuser & Nelly rape accuser* | 81 | .314 | .004 |
| Pair 10 | *Sleep deprivation & inner-city deprivation* | 81 | -.001 | .992 |
| Pair 11 | *Salt intake & normal human intake* | 81 | .609 | .000 |
| Pair 12 | *Police shooting protests & labor protest* | 81 | .395 | .000 |
| Pair 13 | *Tax relief & food relief* | 81 | .356 | .001 |
| Pair 14 | *Seat reservation & computer reservation* | 81 | .186 | .097 |
| Pair 15 | *Star chase & police chase* | 81 | .244 | .028 |

In Table 5, the predictability rate of novel compounds/phrases (2.4188) with OBJECT relation is quite close to that of the old compounds (2.2967) with OBJECT relation. While the predictability rate of the semantic categories of DIRECTION (.8519), MATERIAL (.8272), TOPIC (1.5679) is relatively low.

Statistical result of Table 6 indicates that the priming (Pair 2 *shot block* & *stone block* .366, Pair 3 *drone crackdown* & *military crackdown* .337, Pair 4 shark warning & customer warning .306, Pair 5 *police shooting* & *campus shooting* .397, Pair 9 *Trump accuser* & *Nelly rape accuser* .314, Pair 11 *salt intake* & *normal human intake* .609, Pair 12 *police shooting protests* & *labor protest* .395, Pair 13 *tax relief* & *food relief* .356) and constraining effects (Pair 1 *comet landing* & *practice landing* .219, Pair 6 *bike share* & *market share* .048, Pair 7 *Australia attack* & *bomb attack* .221, Pair 8 *property manager* & *assistant manager* .011, Pair 10 *sleep deprivation* & *inner-city deprivation* -.001, Pair 14 *seat reservation* & *computer reservation* .186, Pair 15 *star chase* & *police chase* .244 ) co-exist between novel compounds/ phrases with the same surface structure.

## 3. DISCUSSION

### 3.1 Syntactic Categories and Predictability

Table 2 indicates that novel synthetic compounds/phrases (O $N_V$ + $N_A$ $N_V$, *M*=36.1358) are closely associated with old synthetic compounds (OV-*ing*, *M*=36.5926 & OV-*ing*, M=32.3086). The old syntactic categories (deep structure) may help to predict novel compounds/phrases. On the other hand, correlations of 6 pairs in Table 3 also reveal that syntactic similarities do not guarantee high rate of predictability because the correlation value of "O $N_V$ + $N_A$ $N_V$ & N $N_V$" (.765) is higher than that of "OV-er & O $N_V$ + $N_A$ $N_V$" (.405). Table 6 shows that priming and constraining effects co-exist between compounds/phrases with the same surface structures.

### 3.2 Semantic Categories and Predictability

Table 4 indicates that syntactic categories and semantic categories do not have a one-to-one matching relationship. Semantic categories (OBJECT, DIRECTION, PROPERTY, MATERIAL, AGENT, TOPIC, LOCATION AT, INSTRUMENT) are far more complicated than syntactic categories (OV-*er*, OV-*ing*, $N_A$ $N_V$, N $N_V$, O $N_V$). Semantic categories of OBJECT (2.4188), PROPERTY (2.4475), AGENT (2.6173), LOCATION AT (2.2037), and INSTRUMENT (2.9568) share similar mean values. While other categories such as DIRECTION (.8519), MATERIAL (.8272), and TOPIC (1.5679) have relatively lower mean values. The mean value of NEW OBJECT (2.4188) is correlated with the OLD OBJECT (2.2967), which means same semantic categories can help to facilitate the predictability of novel compounds/categories. Compared with English speakers, Chinese translators tend to bypass the phrase of syntactic analysis and interpret novel compounds/phrases by semantic processing on the surface structure.

### 3.3 Structural Ambiguity and Interpretation

Under the circumstance of de-contextualization, structural ambiguity is unavoidable. For example, the phrase *shark warning* & *shared-bike hunters* may be generated from two deep structures (OV-*ing* and N $N_V$). Consequently there are two possible competing interpretations ("the warning given to sharks" vs. "the warning given to people about the possible danger of sharks"; "hunters who hunt for and tide up shared-bikes" vs. "hunters who ride on shared-bikes." In these cases, syntactic operations and contextual analysis seem absolutely necessary in meaning interpretation. According to Antonietta Bisetto (2015 academic communication), "the interpretation of a compound depends mainly on your knowledge of the world. […] As for the general case, a N+N compound has a 'preferred' interpretation. This, actually, depends on the language. In English a compound can have more than a single interpretation." Besides syntactic categories and context, world knowledge is another important factor in compound interpretation. Heidi Harley (2015 academic communication) holds similar opinion by declaring "the interpretation 'give warning to sharks' is available for the compound 'shark warning' but is simply so pragmatically unlikely that I am not surprised it is not attested in corpora. You might find examples like 'violator warning' or similar on the interpretation 'a warning to violators.'"

## CONCLUSION

The evidence obtained from the testing experiment proves that syntactic operation and semantic processing are quite important in compound interpretation under de-contextualized condition. Unlike English speakers, Chinese translators often unconsciously neglect the phase of syntactic paraphrase and interpret and re-structure the compounds/phrases by processing surface structures. We assume that syntax knowledge will help Chinese translators to predict and translate English compounds and phrases because sometimes newly emerged compounds/phrases share the same surface structures although they are generated from different deep structures. Translators also need world knowledge, contextual information and pragmatic awareness to interpret and translate some ambiguous compounds/ phrases. Psychological differences and contextual settings shall be considered in further research.

## ACKNOWLEDGMENTS

We are very grateful to Professor Chen Ping, Professor Antonietta Bisetto, Professor Heidi Harley, and Professor

Ray Jackendoff, who provided their valuable opinions via email or academic occasions on the questions mentioned in the current study.

## REFERENCES

Adams, V. (1973). *An introduction to modern English word-formation*. London: Routledge.

Baldwin, T., & Tanaka, T. (2004). *Translation by machine of compound nominals: Getting it right* (pp.24-31). In Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing, Barcelona, Spain.

Bauer, L. (2003). *Introducing linguistic morphology* (2nd ed.). Washington, DC: Georgetown University Press.

Bloomfield, L. (1933). *Language*. Chicago: The University of Chicago Press.

Cao, Y. B., & Li, H. (2002). *Base noun phrase translation using web data and the EM algorithm*. In 19th International Conference on Computational Linguistics, Taipei.

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, *53*(4), 810-842.

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., & Yuret, D. (2009). Classification of semantic relations between nominals. *Language Resources and Evaluation, 43*(2), 105-121.

Girju, R., Moldovan, D. I., Tatu, M., & Antohe, D. (2005). On the semantics of noun compounds. *Computer Speech & Language,* 19(4), 479-496.

Heine, B., Ulrike, C., & Friederike, H. (1991). *Grammaticalization: A conceptual framework*. Chicago: Univeristy of Chicago Press.

Jespersen, O. (1954). *A modern English grammar on historical principles*. London: Geroge Allen & Unwin Ltd.

Lauer, M. (1995). *Designing statistical language learners: Experiments on noun compounds* (PhD thesis). Macquarie University, Australia.

Lauer, M., & Dras, M. (1994). *A probabilistic model of compound nouns*. In Proceedings of the 7th Australian Joint Conference on AI.

Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language, 61*, 30-44.

Liberman, M., & Sproat, R. (1992). The stress and structure of modified noun phrases in English. In I. A. Sag & A. Szabolcsi (Eds.), *Lexical matters* (pp.131-181). CSLI, Stanford University.

Lieber, R., & Štekauer, P. (2009) . Chapter 1. Introduction: Status and definition of compounding. In R. Lieber & P. Štekauer (Ed.), *Oxford Handbook of Compounding* (pp.1-25). Oxford: Oxford University Press.

Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (2nd ed.). München: Beck.

Nastase, V., & Szpakowicz, S. (2003). *Exploring noun-modifier semantic relations* (pp.285-301). In Fifth International Workshop on Computational Semantics (IWCS-5).

Roeper, T., & Siegel, M. E. A. (1978). A lexical transformation for verbal compounds. *Linguistic Inquiry, 9*(2), 199-260.

Rundell, M. (2005). *MacMillan English-Chinese dictionary for advanced learners*. Beijing: Foreign Language Teaching and Research Press.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 42*, 529-567.

Štekauer, P. (2009) . Chapter 14. meaning predictability of novel context-free compounds. In R. Lieber & P. Stekauer (Eds.), *Oxford handbook of compounding* (pp 431-470).

Warren, B. (1978). Semantic patterns of noun-noun compounds. Acta Universitatis Gothoburgensis. *Gothenburg Studies in English, 41,* 1-266.

Xu, Y. H. (2014). The influence of compound word structure on words meaning guessing. *Language Teaching and Linguistic Studies,* (4), 17-23.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9*, 341-368.

## APPENDIX I

### Compound/phrase translation test

NAME CLASS AGE

SECTION A Please translate the following compounds into Chinese

1. Cigarette lighter
2. Shipbuilder
3. Snowblower
4. Bread winner
5. Face-saver
6. Coffee maker
7. Dishwasher
8. Doorkeeper
9. Drought excluder
10. Earth mover
11. Fire-eater
12. Fire-raiser
13. Fire extinguisher
14. Fortune-teller
15. Gasholder
16. Gate-keeper
17. Glassblower
18. Goalkeeper
19. Hairdresser
20. Haidryer

21. Headhunter
22. Housewarming
23. Housekeeping
24. Lawmaking
25. Matchmaking
26. Blood poisoning
27. Peacekeeping
28. Problem-solving
29. Risk-taking
30. Trendsetting
31. Stamp collecting

SECTION B Please translate the following compounds / phrases into Chinese
1. Comet landing
2. Practice landing
3. Shot block
4. Stone block
5. Drone crackdown
6. Military crackdown
7. Shark warning
8. Customer warning
9. Police shooting
10. Campus shooting
11. Bike share
12. Market share
13. Australia attack
14. Bomb attack
15. Property manager
16. Assistant manager
17. Trump accuser
18. Nelly rape accuser
19. Sleep deprivation
20. Inner-city deprivation
21. Salt intake
22. Normal human intake
23. Police shooting protests
24. Labor protest
25. Tax relief
26. Food relief
27. Seat reservation
28. Computer reservation
30. Star chase
31. Police chase

## APPENDIX II

### One-Sample Statistics

| | *N* | Mean | Std. deviation | Std. error mean |
|---|---|---|---|---|
| *Cigarette lighter* | 81 | 3.7037 | 1.01790 | .11310 |
| *Shipbuilder* | 81 | 2.3457 | 1.40677 | .15631 |
| *Snow blower* | 81 | 1.6173 | 1.74333 | .19370 |
| *Bread winner* | 81 | 2.7284 | 1.82354 | .20262 |
| *Face-saver* | 81 | 1.0370 | 1.47855 | .16428 |
| *Coffee maker* | 81 | 3.0247 | 1.58094 | .17566 |
| *Dishwasher* | 81 | 3.5185 | 1.14139 | .12682 |
| *Doorkeeper* | 81 | 3.0864 | 1.55079 | .17231 |
| *Draught excluder* | 81 | .0864 | .50491 | .05610 |
| *Earth mover* | 81 | 1.5802 | 1.82937 | .20326 |
| *Fire-eater* | 81 | .7778 | 1.53297 | .17033 |
| *Fire extinguisher* | 81 | 2.7284 | 1.83038 | .20338 |
| *Fortune-teller* | 81 | 3.1111 | 1.54919 | .17213 |
| *Glassblower* | 81 | .9877 | 1.61628 | .17959 |
| *Goalkeeper* | 81 | 1.9753 | 2.01231 | .22359 |
| *Asset-stripping* | 81 | 1.0123 | 1.42736 | .15860 |
| *Coal Mining* | 81 | 2.8519 | 1.48418 | .16491 |
| *Decision-making* | 81 | 3.6914 | .86084 | .09565 |
| *Housewarming* | 81 | .9506 | 1.52399 | .16933 |
| *Housekeeping* | 81 | .9877 | 1.42736 | .15860 |
| *Job-sharing* | 81 | .4321 | .89356 | .09928 |
| *Lawmaking* | 81 | 3.5556 | .88034 | .09782 |
| *Matchmaking* | 81 | 1.6914 | 1.76523 | .19614 |
| *Blood poisoning* | 81 | 1.7284 | 1.89085 | .21009 |
| *Peacekeeping* | 81 | 3.8272 | .58716 | .06524 |
| *Problem-solving* | 81 | 3.9877 | .11111 | .01235 |
| *Profit-sharing* | 81 | 2.3827 | 1.47959 | .16440 |
| *Risk-taking* | 81 | 3.3333 | 1.25499 | .13944 |
| *Trendsetting* | 81 | 2.2469 | 1.78557 | .19840 |

To be continued

Continued

| | *N* | Mean | Std. deviation | Std. error mean |
|---|---|---|---|---|
| *Stamp collecting* | 81 | 3.9136 | .47952 | .05328 |
| *Comet landing* | 81 | .8519 | 1.24611 | .13846 |
| *Practice landing* | 81 | 1.2963 | 1.63894 | .18210 |
| *Shot block* | 81 | .5309 | 1.28536 | .14282 |
| *Stone block* | 81 | .8272 | 1.12683 | .12520 |
| *Drone crackdown* | 81 | .7037 | 1.10050 | .12228 |
| *Military crackdown* | 81 | 2.0247 | 1.73908 | .19323 |
| *Shark warning* | 81 | 1.8025 | 1.72061 | .19118 |
| *Customer warning* | 81 | 2.0123 | 1.69185 | .18798 |
| *Police shooting* | 81 | 2.8889 | 1.37840 | .15316 |
| *Campus shooting* | 81 | 3.8395 | .76578 | .08509 |
| *Bike share* | 81 | 3.5802 | 1.08241 | .12027 |
| *Market share* | 81 | 3.1235 | 1.47803 | .16423 |
| *Australia attack* | 81 | 1.6420 | 1.49423 | .16603 |
| *Bomb attack* | 81 | 3.6296 | .95452 | .10606 |
| *Property manager* | 81 | 1.7160 | 1.45116 | .16124 |
| *Assistant manager* | 81 | 2.6543 | 1.59813 | .17757 |
| *Trump accuser* | 81 | 1.9877 | 1.66953 | .18550 |
| *Nelly rape accuser* | 81 | 1.3333 | 1.53297 | .17033 |
| *Sleep deprivation* | 81 | 3.5679 | 1.11734 | .12415 |
| *Inner-city deprivation* | 81 | .5679 | 1.24437 | .13826 |
| *Salt intake* | 81 | 3.2222 | 1.40535 | .15615 |
| *Normal human intake* | 81 | 2.8148 | 1.58202 | .17578 |
| *Police shooting protests* | 81 | 2.0864 | 1.71171 | .19019 |
| *Labor protest* | 81 | 2.7531 | 1.52914 | .16990 |
| *Tax relief* | 81 | 3.5185 | 1.15229 | .12803 |
| *Food relief* | 81 | 2.7160 | 1.58299 | .17589 |
| *Seat reservation* | 81 | 3.4198 | 1.22336 | .13593 |
| *Computer reservation* | 81 | 2.2840 | 1.66759 | .18529 |
| *Star chase* | 81 | 3.4568 | 1.37885 | .15321 |
| *Police chase* | 81 | 2.6049 | 1.58650 | .17628 |