STUDIES IN LITERATURE AND LANGUAGE
Vol.**2**, No. **3**, **2011**, pp.30-35
www.cscanada.net

ISSN 1923-1555 [Print]
ISSN 1923-1563 [Online]
www.cscanada.org

Unified Classification of Nominal Classifiers and Formalization of Classifier-noun Phrases*

GUAN Xiaowei[1,†]

HAN Qian[2]

ZHANG Rui[3]

Abstract: One of the main problems that affect the quality of machine translation is how to express the knowledge of language in precision. Based on the theory of Semantic Element (SE) in Unified Linguistics, a new unified classification of English and Chinese nominal classifiers is proposed from the perspective of C-E and E-C translation. Different Semantic Element Representations (SER) of classifiers in English and Chinese have the same semantic type of classifiers. The English and Chinese noun-classifier phrases are formalized into English and Chinese SER respectively.

Key words: Chinese Classifier-Noun Phrase; Classification of Nominal Classifiers; Formalization; SE; SER

DOI: 10.3968/j.sll.1923156320110203.004

INTRODUCTION

The history of machine translation is more than a half century long, but its quality is still poor and the machine translation systems still haven't reached the practical phase. The key lies in linguistics, that is, the knowledge of language cannot be expressed in precision, although many famous linguists have made many important achievements from different angles of study.

Classifier is a class of active and commonly-used words in Chinese, which are seldom used alone, but with numerals to modify nouns, verbs and adjectives, etc. Therefore we should set the study of classifiers into a dynamic context and analyze the relationships between classifiers and their collocated elements. In both linguistics and MT fields, classifiers have been studied mainly in the aspects of classification and semantic collocation. The classification of classifiers will be discussed in Section 2. In MT, Zheng (Zheng, Li, & Chen, 2002) concluded interaction between classifier and its correlate, and illustrated its

***This work is supported by the Fundamental Research Funds for the Central University. Number: DUT10RC(3)46**

[1] Lecturer, Doctor. Main research areas are machine translation and comparison of C-E and E-C translation. School of Foreign Languages, Dalian University of Technology, Dalian, China.
† Corresponding Author. Email: angel_gxw@yahoo.com.cn
[2] School of Foreign Languages, Dalian University of Technology, Dalian, China
[3] School of Foreign Languages, Dalian University of Technology, Dalian, China
‡ Received May 14, 2011; accepted May 25, 2011.

application in semantic disambiguation and structure disambiguation. Zhang (Zhang, 2003) designed a noun-classifier collocation dictionary, and analyzed the classification of classifiers to help parsing semantic and syntactic problems. "The Grammatical Knowledge-base of Contemporary Chinese—A Complete Specification", developed by Institute of Computational Linguistics of Beijing University, made a detailed classification of classifiers and collocated abilities of nouns with classifiers (Yu, Zhu, & Wang, 1998). These studies mainly analyze classifiers in Chinese, seldom relating them with other languages in translation and without analyzing the formalization of classifier-noun phrases.

This paper offers an account of one problem of the precision of natural language: Chinese and English classifier-noun phrases formalization. In section 2, we discuss a new unified classification of English and Chinese nominal classifiers based on the theory of Unified Linguistics. Section 3 shows the formalization of Chinese and English classifier-noun phrases by SE and SER.

1. A UNIFIED CLASSIFICATION OF CHINESE AND ENGLISH CLASSIFIERS

1. 1 Former Study of the Classifications of Classifiers

1. 1.1 Linguistic Classification of Chinese and English Classifiers

Actually, there isn't such a word class as classifier in traditional grammar system in English which belongs to Indo-European Language Family. However, some famous linguists use different terms to describe the concept of "quantity", such as "quantitative determiner", "quantitative and partitive noun", "partitives", "partitive unit noun". Their classifications are also different. Zhang (Zhang, 1998) divides classifiers into five kinds including quantity, shape, capacity, action and pairing, grouping partitive unit nouns. Zhang (Zhang, 2002) doesn't regard classifiers as a word class, but words and phrases which are related with quantity, including pronouns, determiners, adjectives, nouns and compound words related with quantity, numbers, noun-"of" and possessive phrases.

Classifier, which is a class of active and commonly-used words in Chinese, is classified into different kinds according to different criteria and angles. L ü (L ü, 1982) divides classifiers into measures, words borrowed from objects and verbs, units related with groups and time. Zhu (Zhu, 1998) divides classifiers into individual, collecting, measure, indefinite, temporary, verbal and phrase units.

1. 1.2 Classification of Classifiers in NLP

In "The Grammatical Knowledge-base of Contemporary Chinese" (Yu, Zhu, & Wang, 1998), the classifiers fall into three classes: verbal classifier (individual, group, measure, container, shape, times, category, indefinite and compound units), time classifier, and nominal classifier.

HowNet (Dong & Dong) is an online knowledge-base which reveals the relationship among concepts, and the relationship among attributes of concepts. In HowNet, words are described with Concept Definition, and different meanings of a word are defined as different concepts. First Sememe is the smallest semantic unit in a concept. Chinese classifiers are not classified detailedly, but are marked as "CLAS" for some evident classifiers. Many classifiers are classified in other word classes.

1. 2 Unified Classification of Chinese and English Nominal Classifiers

1. 2.1 Classification Theory

The traditional classifications of classifiers, both in linguistics and in NLP, are processed in a certain language, resulting in inconformity of the classifications. According to Unified Linguistics (Gao & Gao, 2009) different language can be translated into each other and people speaking different language can communicate with each other because there are words, phrases and sentences with the same meaning. Natural languages need to be described in a unified way from the perspective of semantics. Semantic of a sentence is called SS. An element to express a semantic meaning in an SS is called Semantic Element

(SE). Semantic language (SL) consists of all SEs, including all SSs. Any natural language can be regarded as a representation of semantic language. The translation between two languages (I, J) is regarded as a transformation between two representations. Therefore, to describe classifiers of different natural languages in a unified way, this paper proposes a unified classification, i.e. classifiers with the same semantic type have the same SE. The translation of classifiers between Chinese and English is the transformation between two different representations in Chinese and English.

1. 2.2 Unified Classification of Nominal Classifiers

Chinese and English classifiers are classified in a unified criterion from bi-lingual angle in the interest of C-E and E-C machine translation.

Classifiers are classified into eight kinds, including Formalized Classifier, Group Classifier TypeⅠ and Ⅱ, Precise Measure Classifier, Container Classifier, Pair & Group Classifier, Shape Classifier, Category Classifier.

(1) Formalized Classifier (FC) is the unit of nouns representing things, and a kind of abstract and formalized Chinese characters without real meaning. FC is unique in Chinese.

In English, when we want to represent "a single thing", we often use an indefinite article plus a singular countable noun. But in Chinese, because of the habitual expression forms of the language, to express this kind of semantic meaning, we must use the form "number + classifier+singular countable noun. For example, the Chinese translation of English phrase "a man" is "一個男人 *yi ge nanren*". Here, "個 *ge*" is classifier, "男人 *nanren*" means "man". In E-C MT, for this kind of collocation, the indefinite articles often need to be translated, except for some special cases that they don't need to. According to the unique expression practice of Chinese, a proper classifier should be added between a numeral and a noun to express a complete meaning. Moreover the added classifier has close semantic relation with the collocated noun.

However, we must distinguish this kind of collocation phenomenon with another kind of collocation with similar surface structure. For example, the corresponding Chinese translation of the English pattern "a pair of N" is "一雙 N (*yi shuang* N)". Here, "雙 *shuang*" means "pair". Although "*shuang*" is also a classifier, it is different from "*ge*" which we discussed just now. Because we can find a corresponding translation of "*shuang*" in English or other languages, but we can't find a corresponding translation of "個 *ge*" in English or other languages. So this kind of classifier like "*ge*" is just a formalized representation with no real semantic meaning. They are just a kind of symbols in Chinese representation.

(2) Group Classifier. For C-E translation, we divide traditional group classifiers into two types. Group Classifier TypeⅠrefers to the group classifiers selected according to the semantics of the collocated nouns. Group Classifier TypeⅡrefers to the group classifiers excluding TypeⅠ.

In many languages, we may have observed such a common phenomenon, that is, we often use some quantifiers to express the meaning of "a large number of something which gathered together". In Chinese we use the pattern "一群 N (*yi quan* N)" to represent this semantic meaning. Here, "一 *yi*" denotes "one", which is a numeral, "群 *quan*" denotes the representation of the semantic meaning of "large number". In Chinese there exists no such kind of language phenomenon, but in some western languages this phenomenon is very common, such as in English, French and German.

(3) The rest five kinds of classifiers are almost the same as defined by other scholars. In such classifier-noun phrases, we can find the corresponding Chinese and English classifier words.

Table 1 shows the Unified Classification of Chinese and English nominal classifiers and their examples.

Table 1: Unified Classification of Chinese and English Nominal Classifiers

Classifier Type	*Chinese example*	English example
Formalized Classifier(for Chinese only)	匹 *pi*/只 *zhi*	——
Group Classifier TypeⅠ	群 *qun*	crowd/herd/cluster
Group Classifier TypeⅡ	幫 *bang*/夥 *huo*	bunch/gang
Precise Measure Classifier	克 *ke*	gram
Container Classifier	瓶 *ping*	bottle
Pair & Group Classifier	雙 *shuang* 對 *dui*/付 *fu*	pair
Shape Classifier	張 *zhang*/件 *jian*/塊 *kuai*	piece
Category Classifier	種 *zhong*	kind

2. THE FORMALIZATION OF CHINESE AND ENGLISH CLASSIFIER-NOUN PHRASES

2. 1 The Theory of SE and SER

In 1980's, Institution of Computing Technology (ICT) of the Chinese Academy of Science (CAS) proposed a new concept (Gao, Chen, & Li, 1989), i.e., semantic element representation (SER) with variables and without variables. The concept of Semantic Element and Semantic Language was discussed in detail in Gao, Q. S., & Gao, X. Y. (2009)[4] and Gao, Q. S., Hu, Y., & Gao, X. Y. (2003)[5].

An element to express a semantic meaning in an SS is called Semantic Element (SE). The representation of an SE in a natural language-I, such as English, Chinese…, is called the Representation of Semantic Element in Language-I (SERi). Semantic of SER is SE. For example, the Chinese sentence is "李先生是教授 *Li xiansheng shi jiaoshou* (Mr. Li is a professor.)". The four SEs in this sentence are 李(Li)，先生($X_{surname}$) (Mr. (X))，教授 (professor)，是 $_{title}(X_{people},Y_{title})$ ($Is_{title}(X_{people},Y_{title})$). "X" and "Y" are two parameters. SE is an abstract concept and form. Actually, the above mentioned examples of SEs are only some remembrance forms to represent SE by using Chinese characters or English words. We can use "1" to substitute the SE "Li", or use "4(X_{people},Y_{title})" or "4(N_{people},N_{title})" to substitute "是 $_{title}(X_{people},Y_{title})$". The subscript denotes the semantic type of the word. A SER can reflect the semantics and syntax relations of the word in a phrase or a sentence.

2. 2 The Formalization of Chinese and English Classifier-Noun Phrases

Chinese and English classifier-noun phrases are formalized based on the theory of SE and SER. SERc denotes Chinese SER; SERe denotes English SER. In classifier-noun phrase, the meaning of noun is defined as Thing Semantic, which is the Semantic meaning expressed by nouns representing things. The collocated nouns are described as "N_x". Here, x denotes the semantic type of the noun. Nouns with the same semantic type can be substituted by each other in a SER.

We use "L_{FC}" to represent FC. In English, besides indefinite-article-noun collocation, collocations such as numeral-noun, ordinal-noun and determiner-noun also need to be added proper FC while being translated into Chinese. These determiners include "this, that, each, every, both, one of, next, same, another, the other, several, a few, etc". The SERe of a FC-noun phrase is "a/Num/this/… N", the corresponding SERc is "一 *yi*/Num/this/… L_{FC} (N) N". "L_{FC} (N)" shows there exists a function relation between L_{FC} and noun, that is, the selection of L_{FC} depends on the specific collocated noun form.

For Group Classifier TypeⅠ and noun phrase, we only extract the noun as parameter, the group classifier and other words in the phrase should remain both in SERe and SERc.

[4] Gao, Q. S., & Gao, X. Y. (2009). *Foundations of Unified Linguistics*. Beijing: Science Press.
[5] Gao, Q. S., Hu, Y., & Gao, X. Y. (2003). Semantic Language and Multi-Language MT Approach Based on SL. *Journal of Computer Science and Technology, 18*, 848-852.

For the rest six kinds of classifiers, the classifier and noun in a classifier-noun phrase are extracted as parameters in the SER., other words in the phrase should remain.

Table 2 shows the examples of SERe and SERc of classifier-noun phrases for each type of classifiers.

Table 2: Examples of SERe and SERc of Classifier-noun Phrases

Classifier Type	English example	Chinese example	SERe	SERc
Formalized Classifier	a hen three trees	一隻母雞(*yi zhi muji*) 三棵樹(*san ke shu*)	a N ——	一 *yi*/Num L_{FC} (N) N L_{FC} (N)=只 *zhi*（N=母雞 *muji*|…） L_{FC} (N)=棵 *ke*（N=樹 *shu*|…）
Group Classifier Type Ⅰ	a herd of elephant	一群象(*yi qun xiang*)	a herd of N_{beast}	一群 *qun* N_{beast}
Group Classifier Type Ⅱ	a bunch of hooligans	一幫流氓 (*yi bang liumang*)	a $L_{ⅡGC}$ of $N_{bad\ person}$	一 $L_{ⅡGC}$ $N_{bad\ person}$
Precise Measure Classifier	a kilo of oranges	一公斤桔子 (*yi gongji juzi*)	a L_{PMC} of N_{fruit}	一 L_{PMC} N_{fruit}
Container Classifier	a cup of coffee	一杯咖啡(*yi bei kafei*)	a L_{CC} of N_{drink}	一 $L_{容器}$ $N_{饮品}$
Pair & Group Classifier	a pair of shoes	一雙鞋(*yi shuang xie*)	a L_{PGC} of N_{shoe}	一 L_{PGC} N_{shoe}
Shape Classifier	a piece of paper	一張紙(*yi zhang zhi*)	a L_{SC} of N_{paper}	一 L_{SC} N_{paper}
Category Classifier	a kind of birds	一種鳥(*yi zhong niao*)	a L_{CaC} of N_{animal}	一 L_{CaC} N_{animal}

CONCLUSION

This paper focuses on solving one of key problems of describing language phenomenon precisely in MT—the classification of classifiers and formalization of classifier-nouns phrases in Chinese and English. A unified classification is proposed based on Unified Linguistics Theory and the classifier-nouns phrases are formalized by SE and SER. The study in the paper will beneficial to the process of word selection in MT. More efforts should be made to put this method into practice, and to improve the quality of lexical translation in machine translation.

REFERENCES

Zheng, X. L. , Li, T. Q., & Chen, Y. D. (2002). Application of Interaction between Classifier and its Correlate in Chinese Phrase Disambiguation. *Journal of Xiamen University, 41*, 715-719.

Zhang, H. (2003). *Building noun-classifier collocation dictionary oriented to chinese information processing*, MSc thesis, Shanghai Jiao Tong University.

Yu, S. W., Zhu, X. F., & Wang, H. (1998). *The grammatical knowledge-base of contemporary chinese - a complete specification*. Beijing: Tsinghua University Press.

Zhang, Z. B. (1998). *New english grammar course*. Shanghai: Shanghai Foreign Language Education Press.

Zhang. D. Z. (2002). *A practical english grammar*. Beijing: Foreign English Teaching and Research Press.

L ü, S. X. (1982). *Essentials of Chinese grammar*. Beijing: The Commercial Press.

Zhu, D. X. (1998). *Grammar note*. Beijing: The Commercial Press.

Dong, Z. D., & Dong, Q.. HowNet. Available: www.keenage.com

Gao, Q. S., & Gao, X. Y. (2009). *Foundations of unified linguistics*. Beijing: Science Press.

Gao, Q. S., Chen, Z. X., & Li, T. Q. (1989). The principle of human-like machine translation. *Computer Research & Development, 26*, 1-7.

Gao, Q. S., Hu, Y., & Gao, X. Y. (2003). Semantic language and multi-language mt approach based on sl. *Journal of Computer Science and Technology, 18*, 848-852.