Two-level Word Class Categorization Model in Analytic Languages and Its Implications for POS Tagging in Modern Chinese Corpora

Renqiang Wang
Graduate School,
Sichuan International Studies University,
Chongqing 40031
wangrenqiang@sisu.edu.cn

Changning Huang
Department of Computer Science and Technology, Tsinghua University,
Beijing 100084
cnhuang0908@126.com

Abstract

The study of word classes has a history of over 4000 years, and the word class problem in over 1000 analytic languages like Modern Chinese can be seen as the Goldbach Conjecture in linguistics. This paper first outlines the existing problems in the POS tagging of Modern Chinese corpora with a case study of 自信. Then it introduces the Two-level Word Class Categorization Model in analytic languages, which is based on the perspectives of language as a complex adaptive system and the nature of major parts of speech as propositional speech act functions. Finally, the implications of Two-level Word Class Categorization Model for POS tagging in Modern Chinese corpora are explored.

1 Introduction

Categorization is a fundamental task in linguistics, and linguistic categories like word classes or parts of speech were considered as the study of "god particles" in language in the 36th Annual Conference of the German Linguistic Society held at the University of Marburg, Germany, in March, 2014. In natural language processing, part-of-speech tagging plays a key role. As pointed out by Rabbi (2012), "It is a significant pre-requisite for putting a human language on the engineering track." The study of word classes has a history of over 4000 years, but the word class problem in over one thousand analytic languages like Modern Chinese, Modern English and Tongan can be seen as the Goldbach Conjecture in linguistics, which has witnessed several upsurges over the last century.

Let's take the example of 自信 in Chinese. The first five editions of *The Contemporary Chinese Dictionary* (hereinafter called CCD) have almost the same treatment of 自信 with the only definition of 相信自己, which is obviously a verbal usage according to the definition metalanguage, though it is only in CCD5 published in 2005 that the lexeme is explicitly labeled as VERB:

【自信】 zìxìn 动 相信自己：～心｜～能够完成这个任务。

In CCD6 published in 2012, however,

29th Pacific Asia Conference on Language, Information and Computation pages 1 - 10
Shanghai, China, October 30 - November 1, 2015

we can see that 自信 is labeled as a multi-category lexeme belonging to VERB, NOUN and ADJECTIVE:

> 【自信】zìxìn ❶ 动 相信自己：～心 | ～能够完成这个任务。❷ 名 对自己的信心：不能失去～ | 工作了几年之后，他更多了几分～。❸ 形 对自己有信心：他做事总是很～。

In the second edition of *The Grammatical knowledge-base of Contemporary Chinese — A Complete Specification* (Yu et al., 2003), 自信 is specified only as VERB with the following examples, which illustrate its typical usages:

> ～心 | 他～自己能考取北京大学/我～能完成任务/～地说/在困难面前，需要～

Then what about the POS tagging of 自信 in Chinese corpora? We downloaded all the concordance lines from the Modern Chinese Corpus developed by the China National Language and Character Working Committee (hereinafter called CN CORPUS, http:// cncorpus.org/ CCindex. aspx). There are altogether 187 downloadable concordance lines of 自信. As shown in Table 1, the most frequent usages of 自信 are as VERB and ADJECTIVE, with only one occurrence as NOUN.

	parts of speech	frequency	percentage
1	VV	142	75.94%
2	JJ	43	22.99%
3	NN	1	0.53%
4	word-formation morpheme	1	0.53%
total		187	100.00%

Table 1: POS Tagging of 自信 in CN CORPUS

However, through careful analysis, we find that 117 of them (accounting for 62.54%) seem to have problems in their POS tagging. Though the usages of 自信 in the corpus are respectively tagged as VERB, ADJECTIVE and NOUN, which seems to be consistent with the word class labeling in CCD6, we have found the following five types of problematic POS tagging in CN CORPUS:

First, usages of reference when used as subjects or objects of the sentences are tagged differently with the parts of speech of NOUN as in (1), ADJECTIVE as in (2), (3), (8), (9) and (12), and VERB as in (4), (5), (6), (7), (10), (11), (13), (14) and (15). Admittedly, not all of them are correct tagging. Moreover, usages of 自信 classified by 一种 are all tagged as VERB as in (5), (6) and (7), which are typical nominal usages. Interestingly, juxtaposed words as objects of the sentences are obviously NOUN like 激情 and 力量 while 自信 are still tagged as VERB, as in (11) and (12).

(1) 话/n 虽/c 这么说/v ，/w 织云/nh 也/d 并/c 没有/v 多少/m 自信/n

(2) /w 声音/n 里/nd 没有/v 一点/m 自信/a ，/w 连/p 她/r 自己/r 也/d 感觉/v 到了/v 。/w

(3) 他/r 那/r 种/q 到/v 哪儿/r 、/w 永远/d 吃/v 得/u 开/v 的/u 自信/a 从/p 何/r 而/c 来/vd ？/w

(4) 聪明/a 、/w 好学/v 、/w 自信/v 是/vl 王惠莹/nh 的/u 突出/a 特点/n 。/w

(5) w 在/p 中国/ns 模特/n 身上/nl 有/v 一/m 种/q 发自/v 内心/n 的/u 自信/v …/w …/w

(6) 他/r 笑/v 了/u ，/w 眸子/n 里/nd 透出/v 一/m 种/q 自信/v 。/w

(7) 但/c 她/r 时时/d 表现/v 出/vd 一/m 种/q 能/vu 战胜/v 危险/a 的/u 自信/v 。/w

(8) 雷嘉/nh 帮助/v 她/r 获得/v 了/u 冷静/a 和/c 自信/a 。/w

(9) 他/r 充满/v 了/u 对/p 自己/r 这/r 一代/nt 人/n 的/u 骄傲/a 和/c 自信/a 。/w

(10) 口气/n 充满/v 了/u 自信/v 。/w

(11) 一/m 个/q 人/n 只要/vu 真正/a 树立/v 了/u 对/p 祖国/n 、/w 对/p 人民/n 、/w 对/p 社会/n 的/u 责任感/n ，/w 就/d 会/vu 自觉/a 地/u 对/p 生活/n 充满/v 激情/n 和/c 自信/v

(12) 他/r 又/d 恢复/v 了/u 自信/a 和/c 力量/n 。/w

(13) /w 他/r 怔怔/a 地/u 看/v 着/u 我/r ，/w 但/c 很快/a 又/d 恢复/v 了/u 自信/v

(14) 她/r 的/u 笑容/n 中/nd 蕴含/v 着/u 对/p 改革/v 的/u 无限/d 自信/v 。/w

(15) 我/r 仔细/a 想/v 着/u ，/w 把/p 花/n 角儿/n 的/u 动作/n 合理化/v ，/w 使/v 自己/r 增加/v 自信/v

Secondly, usages of modification of entities are tagged differently with the parts of speech of ADJECTIVE as in (16) to (18), and VERB as in (19) to (24) , even when juxtaposed words like 平静, 刚愎, 愉快 and 自大 are tagged as ADJECTIVE as in (19), (21), (22) and (23).

(16) /w 吉明/nhs 本来/d 是/vl 个/q 坚强/a 自信/a 的/u 青年/n

(17) /w 她/r 的/u 眼睛/n 不如/v 水子/nh 灵气/n ， /w 透出/v 刚毅/a 自信/a 的/u 光芒/n ；

(18) 渐渐/a 他/r 的/u 脸色/n 恢复/v 了/u 常态/n ， /w 又/d 浮上/vd 了/u 他/r 平日/n 那/r 种/q 自信/a 和/c 冷漠/a 的/u 神气/n

(19) /w 一个/mq 平静/a 而/c 又/d 自信/v 的/u 声音/n ，/w 在/p 我们/r 身后/nl 响起/v 。

(20) /w 当/p 他/r 年轻/a 的/u 时候/n 他/r 是/vl 非常/d 自信/v 的/u 人/n 。/w

(21) 刚才/d 还/d 刚愎/a 自信/v 的/u 斐烈/nh ，/w 这时候/nt 抓耳挠腮/i ，/w 无可奈何/i 地/u 摇/v 了/u 摇头/v 。/w

(22) 他/r 那/r 愉快/a 的/u 自信/v 的/u 调子/n ，/w 好象是/v 他/r 在/p 指挥/v 着/u 它们/r 似的/u 。

(23) 在/p 他们/r 这/r 种/q 自信/v 的/u 心理/n ，/w 也/d 可/vu 说/v 是/vl 自大/a 的/u 心理/n ，/w 这/r 种/q 精神/n 胜利/v 便/d 成为/v 绝对/a 不可/vu 缺/v 之/u 物/n 。/w

(24) 看/v 着/u 王惠莹/nh 领奖/v 时/nt 自信/v 的/u 面容/n ， /w 许多/a 体操/n 行家/n 和/c 新闻记者/n 都/d 预言/v

Thirdly, usages of predicative adjectives of 自信 are tagged differently: some are tagged as ADJECTIVE as in (25) to (28), whereas others as VERB as in (29) to (33), even when juxtaposed words like 精干, 果断 and 平静 are tagged as ADJECTIVE as in (30), (32) and (33).

(25) /w 他/r 的/u 口气/n 倔强/a 而/c 自信/a 。/w

(26) 中国人/n 哟/u ，/w 是/vl 大胆/a 、/w 自信/a 的/u ， /w 有时/d 甚至/d 是/vl 执拗/a 的/u 。

(27) /w 他/r 的/u 神色/n 显得/v 更/d 庄严/a 、/w 更/d 高傲/a 和/c 更/d 自信/a 了/u 。/w

(28) 但/c 那时/nt 你/r 年轻/a ， /w 自信/a ， /w 浑身/n 洋溢/v 着/u 青春/n 的/u 活力/n

(29) 也许/d 他/r 太/d 过于/d 自信/v ， /w 命运/n 竟/d 捉弄/v 了/u 他/r —/w

(30) /w 模样/n 儿/k 很/d 精干/a ，/w 也/d 很/d 自信/v 。

(31) 灰灰/nh 可/d 自信/v 啦/u ，/w 他/r 说/v ： /w "/w 是/vl 红海/ns ！/w

(32) 白脖黑/n 她/r 从来/d 都/d 是/vl 走/v 在/p 鸭/n 群/n 队伍/n 的/u 第/h 一个/mq ：/w 挺/v 着/u 胸脯/n ，/w 自信/v 而/c 又/d 果断/a

(33) 王/nhf 所长/n 用/p 疑惑/v 的/u 目光/n 望/v 着/u 冯/nhf 教授/n ，/w 教授/n 还是/d 那么/r 平静/a 而/c 自信/v ： /w

Fourthly, usages of 自信 plus 地 in adverbial constructions are tagged differently: ADJECTIVE in (34) and (35) while VERB in (36) to (38).

(34) 戈华/nh 非常/d 自信/a 地/u 判断/v 说/v 。

(35) 高福源/nh 很/d 自信/a 地/u 表示/v ： /w "/w 我/r 自己/r 既然/c 要求/v 回去/v ，/w 就/d 有/v 这个/r 把握/v 。/w

(36) 黑仔/nh 作/v 了/u 一下/mq 深呼吸/v ， /w 十分/d 自信/v 地/u 说/v 。

(37) 盟军/n 总参谋长/n 自信/v 地/u 用/p 指示/n 棍/n 指点/v 着/u 墙上/nl 的/u 军用地图/n

(38) 我/r 自信/v 地/u 说/v ：/w "/w 我/r 要/vu 发明/v 一/m 种/q 更/d 理想/a 的/u 东西/n ， /w 是/vl 给/p 人/n 吃/v 的/u ！/w

Lastly, word-formation usages of 自信 are tagged differently: no tagging in (39) while VERB in (40) to (42), the latter of which seems somewhat awkward .

(39) /w 提高/v 全/a 民族/n 的/u 自信心/n 更有/v 其/r 伟大/a 意义/n 。/w

(40) 鲁迅/nh 先生/n 曾/d 对/p “/w 不/d 失掉/v 自信/v 力/n 的/u 中国人/n ”/w 给予/v 热烈/a 的/u 赞颂/v

(41) 一个/mq 国家/n ，/w 一个/mq 民族/n ，/w 如果/c 没有/v 自信/v 力/n 就/d 不/d 可能/vu 振兴/v 社稷/n

(42) /w 使/v 人/n 在/p 认知/v 上/nd 建立/v 了/u 极/d 大/a 的/u 安全感/n 与/c 稳定/a 感/n ， /w 以及/c 对/p 自己/r 的/u 自信/v 感/n 。

To sum up, we have the following questions: (1) Both the first five editions of CCD and the second edition of *The Grammatical knowledge-base of Contemporary Chinese — A Complete Specification* seem to have adhered to the Principle of Parsimony (namely fewest possible multi-category words), as advocated by Lü Shuxiang (1979), Zhu Dexi (1985), Guo Rui (2002), Lu Jianming (1994, 2013), Yu Shiwen et al (2003) and Shen Jiaxuan (2009, 2012), but then is the word class labeling of 自信 as VERB, NOUN and ADJECTIVE in CCD6 correct? (2) What's the relationship between the word class labeling of lexemes in Chinese dictionaries and the part-of-speech tagging in Chinese corpora? (3) How to improve the part-of-speech tagging in Chinese corpora? To properly answer the above questions, we will first introduce the Two-level Word Class Categorization Model (TLWCCM) in analytic languages and then discuss its implications for the part-of-speech tagging in Chinese corpora.

2 Two-level Word Class Categorization Model (TLWCCM)

2.1 The Theoretical Model

The multifunctionality / heterosemy / multiple class membership of lexemes in many languages has remained a contentious issue ever since linguistics emerged as an independent discipline in the 19th century. And van Lier & Rijkhoff (2013: 1) considers it as "[c]urrently one of the most controversial topics in

linguistic typology and grammatical theory".

Based on the perspectives of language as a complex adaptive system (Beckner et al, 2009; Larsen-Freeman & Cameron, 2008; Lee et al, 2009; Bybee, 2010) and the nature of major parts of speech as propositional speech act functions proposed by Croft (1991, 2001) and Croft & van Lier (2012) on the basis of Searle (1969), Wang (2014a) argues in his Two-level Word Class Categorization Model in Analytic Languages that just as there are two states of existence of word at the two levels of *langue* (i.e. word type or lexeme in lexicon in a communal language) and *parole* (i.e. word token in syntax), word class categorization also happens at the two levels: (1) The word token categorization in syntax at *parole* is the speaker's expression of propositional speech act functions like reference, predication and modification, whereas the word type categorization in lexicon at *langue* is the conventionalized propositional speech act function(s) of a word type resulted from self-organization or collective unconscious; (2) The class membership of a word type does not have *a priori* existence, nor is it precategorial, but is liable to change through recurrent use in various propositional speech act constructions in syntax at *parole*; (3) The multifunctionality or multiple class membership of word types in synchrony derives from diachronic change and is closely related to frequency of use, which reveals the competing motivations of economy and iconicity in communication; (4) The class membership (either single or multiple class membership) of a word type is its meaning potential(s) at *langue*, which is to be discovered by descriptive linguists through corpus-based usage pattern surveys, as is done by dictionary compilers in word class labeling, whereas the class membership of a word token is its meaning as an event as expressed in a certain context, which normally has a single part of speech; (5) With regard to the class membership of a word token, there are prototypical correlations between propositional speech act functions and semantic classes.

2.2 Empirical Studies

Four empirical studies have been conducted with regard to the Two-level Word Class Categorization Model.

Wang (2013) surveys the multiple class membership in Modern Chinese based on CCD5. It is found that 2778 lexemes (accounting for 5.40%) in CCD5 are multi-category words, that multiple class membership exists typically between the major word classes of NOUN, VERB, ADJECTIVE and ADVERB, and that CCD5 has basically labeled with more accuracy the typical parts-of-speech for the headwords and the typical members of the relevant word classes but it is somewhat conservative in dealing with multiple class membership. More importantly, the description of the headwords in the dictionary is partially consistent with the reality of language use in the Chinese community, which reveals the invalid theoretical basis for multiple class membership: the so-called "Principle of Simplicity" in grammar analysis which sticks to the principle of "fewest possible multi-category words" is proved to be problematic.

Wang (2014b) investigates the current status of multiple class membership in Modern English based on *Oxford Advanced Learner's English Dictionary* (7th ed.) (hereinafter called OALD7). It

has been found that 4861 lexemes (accounting for10.48%) in OALD7 are multi-category words, that there are 81 different types of multiple class membership, the most typical of which are those between the major word classes of NOUN, VERB, ADJECTIVE and ADVERB, and that multiple class membership is characteristic of analytic languages like Modern English and Modern Chinese in lexicon at *langue*. Interestingly, the types of multiple class membership in Modern Chinese is similar to that of Modern English, though CCD5 minimized the number of multi-category lexemes by following the Principle of Parsimony/Simplicity, creating a false impression that the percentage of multi-category lexemes in Modern Chinese is far lower than that in Modern English. It is found that this false impression results to some degree from the ban of multiple class membership especially for self-reference lexemes advocated by leading scholars like Zhu (1985), Guo (2002), and Shen (2009), who argue for multifuntionality of Chinese word classes rather than Chinese lexemes. However, this has obviously led to indeterminacy of Chinese word classes.

Wang & Chen (2014) makes a corpus-based study of the relationship between verbs and constructions and proposes four criteria to measure conventionalization of a word's usage (namely, type frequency, token frequency, time span and register variation). It is shown that lexicon and syntax form a continuum with two ends, and that the relationship between verbs and constructions is interdependent in that the verb itself is liable to change through repetitive use in constructions. It is found that the erroneous conclusions in previous studies result from not adopting the corpus-based bottom-up approach, leading to the difficulty of distinguishing the class membership of word types in lexicon at *langue* and that of word tokens in syntax at *parole*, and from committing the logical fallacy of overgeneralization.

Wang & Zhou (2015) makes an empirical study of the correlation between multiple class membership and frequency on the basis of the CN CORPUS and the DIY Word Class Labeling Database of CCD5. The findings of both studies have verified the positive correlation between heterosemy and frequency, but there is a significant difference between them. It is found that the correlation between heterosemy and frequency in analytic languages like Modern Chinese and Modern English results from the competing motivations of economy and iconicity in communication, and that CCD5 minimized the number of multi-category lexemes by following the Principle of Parsimony, creating a false impression that the percentage of heterosemy in Modern Chinese is far lower than that in Modern English.

3 Implications of TLWCCM for POS Tagging in Modern Chinese Corpora

Part-of-speech tagging is the process of assigning a part of speech to each word token in a corpus. From the perspective of TLWCCM, POS tagging is the word class categorization at the level of *parole* in syntax, in which propositional speech act functions (i.e., reference, predication and modification) correlate in markedness patterns with semantic types (i.e., objects, actions , and properties) in contexts.

Accordingly, we can make some

corrections in the above problematic POS tagging in CN CORPUS: 自信 in concordances lines like (12) "他/r 又/d 恢复/v 了/u 自信/a 和/c 力量/n " should be retagged as NOUN instead of ADJECTIVE; 自信 in concordances lines like (13) "他/r 怔怔/a 地/u 看/v 着/u 我/r ，/w 但/c 很快/a 又/d 恢复/v 了/u 自信/v " should be retagged as NOUN instead of VERB; 自信 in concordances lines like (19) "一个/mq 平静/a 而/c 又/d 自信/v 的/u 声音/n ，/w 在/p 我们/r 身后/nl 响起/v " should be retagged as ADJECTIVE instead of VERB; and 自信 in concordances lines like (30) "/w 模样/n 儿/k 很/d 精干/a ，/w 也/d 很/d 自信/v " should be retagged as ADJECTIVE (i.e. predicative adjective) instead of VERB.

Thus, multi-category lexemes like 自信 can cause tag ambiguity in POS tagging in corpora. But how hard is the tagging problem? Or how common is tag ambiguity? Jurafsky & Martin (2009: 135) describes the situation in English:

> It turns out that most words in English are umambiguious; that is, they have only a single tag. But many of the most common words in English are ambiguous..... In fact, DeRose (1988) reports that while only 11.5% of English word types in the Brown corpus are ambiguous, over 40% of Brown tokens are ambiguous.

From the perspective of TLWCCM, tag ambiguity in POS tagging can be removed easily in context (namely in syntax at *parole*). As pointed out in Section 1, many leading scholars in Chinese grammar and Chinese natural language processing adhere to the Principle of Parsimony so as to minimize the scope of multiple class membership or tag ambiguity, and instead argue for multifunctionality of word classes rather than that of lexemes, which is theoretically invalid and practically unnecessary. As verified by Wang Renqiang & Zhou Yu (2015), there is positive correlation between heterosemy and frequency in Modern Chinese. Harbsmeier (1998: 138) correctly pointed out that, in English as in Chinese, the context “painlessly removes the ambiguity of constructions which, taken in isolation, would have been ambiguous”.

This observation has its positive effects on POS tagging in Modern Chinese corpora. According to Bakeoff (2008), among the 5 POS tagged corpora in the survey, 3 are based on the word class information in dictionaries while 2 are token-based. Huang and Huang (2014) found out that the machine learnability of the latter 2 corpora is 2-4 percent higher than the former 3, which indicates that the accuracy of automatic POS tagging can be improved dramatically if we tag the class membership of word tokens in syntax.

Now, if we retag all the problematic concordance lines of 自信 from CN CORPUS from the perspective of TLWCCM, we can get the following results as shown in Table 2. Compared with the original results in Table 1, the number of nominal tags of 自信 has risen dramatically while the number of verbal tags of 自信 has dropped sharply. From Table 2, we can also reach a conclusion that the verbal, nominal and adjectival usages of 自信 are conventionalized, and that CCD6 is right to label 自信 as a multi-category lexeme belonging to VERB, NOUN and ADJECTIVE. According to *Lexicon of Common Words in Contemporary Chinese* released by the

China National Language and Character Working Committee in 2008, 自信 is ranked 3904, which implies that 自信 is a relatively higher frequency lexeme. This obviously explains why it has a higher chance to become a multi-category lexeme and why the accuracy rate POS tagging of 自信 is so low in CN CORPUS.

	parts of speech	frequency	percentage
1	VV	30	16.04%
2	JJ	84	44.92%
3	NN	64	34.22%
4	word-formation morpheme	9	4.81%
total		187	100.00%

Table 2: Results of Revised POS Tagging of 自信

It must be admitted that compared with CCD5, some improvements have been made in CCD6 with regard to word class labeling, but not so much. Our recent survey reveals that for many of the most common words, similar problems still remain: The Principle of Parsimony is still blindly followed. For example, there are still problems in CCD6 in treating lexemes like 研究,方便, 男性, 女性, 自燃, 自杀, 他杀, 拔河, 滑雪, 突变, 渐变, and so on. That's why Huang & Jin (2013: 187) maintains the criteria of POS tagging based on X-Bar Theory, which is to some extent similar to TLWCCM with regard to the word class categorization in syntax at *parole*. And that's also why Huang & Wang (2015) argues that lifting the ban on self-reference senses of multi-category words is an important way out of the Chinese word class dilemma. Since many tagging algorithms require a dictionary that lists all the conventionalized parts-of-speech of every lexeme (Jurafsky & Martin, 2009: 160), the problem now is not that dictionaries are not helpful in POS tagging in analytic languages like Modern Chinese, but that current Chinese dictionaries like the authoritative CCD6 are yet to be the reliable basis for POS tagging in Modern Chinese corpora.

4 Conclusion

To summarize, there is urgent need to improve both the word class labeling in Chinese dictionaries and the POS tagging in Chinese corpora, in which the former often serves as the basis for the latter. And the Two-level Word Class Categorization Model has proved to be effective in providing the guidance for both.

References

Changning Huang and Guangjin Jin. 2013. Three problems of Chinese grammar observed from the Penn Chinese Treebank. *Language Sciences*, (2): 178-192.

Changning Huang and Renqiang Wang. 2015. Lifting the ban on self-reference senses of multi-category words is an important way out of the Chinese word class dilemma. In Proceedings of The 16th Chinese Lexical Semantic Workshop, held in May 2015 at Beijing Normal University.

Changning Huang and Wanmei Huang. 2014. Learnability – A quantititative index of comparative tagsets. *Proceedings of the International Conference on Chinese Word Classes*, Central China Normal University, Wuhan, October 10-10-11.

Christoph Harbsmeier. 1998. *Science and Civilization in China, Volume 7, Part I: Language and Logic*. Cambridge University Press, Cambridge, UK.

Clay Beckner, et al. 2009. Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1): 1-26.

Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Second Edition)*. Pearson Education Inc.

Deborah A. Coughlin. 1996. Deriving part of speech probabilities from a machine-readable dictionary. In *Proceedings of the Second International Conference on New Methods in Natural Processing*. Ankara, Turkey: 37-44.

Dexi Zhu. 1985. *The Questions and Answers on Grammar*. The Commercial Press, Beijing, China.

Diane Larsen-Freeman and Lynne Cameron. 2008. *Complex Systems and Applied Linguistics*. Oxford University Press, Oxford.

Guangjin Jin and Xiao Chen. 2008. The Fourth International Chinese Language Processing Bakeoff : Chinese Word Segmentation, Named Entity Recognition and Chinese POS Tagging. In *Proceedings of SIGHAN-2008*, Vol. 1 (pp.69-81). Hyderabad, India, January 8-10.

Ihsan Rabbi. 2012. Part of Speech Tagging for Pashto. LAP LAMBERT Academic Publishing.

Jan Rijkhoff and Eva van Lier. 2013. *Flexible Word Classes: A Typological Study of Underspecified Parts-of-speech*. Oxford: Oxford University Press.

Jianming Lu. 2013. *A Course Book of Modern Chinese Grammar Research (4th edition)*. Peking University Press, Beijing, China.

Jiaxuan Shen. 2009. My view of word classes in Chinese. *Language Sciences*, (1): 1-12.

Jiaxuan Shen. 2012. Reflections on "nouny verbs": Problems and solutions. *Chinese Teaching in the World*, (1): 3-17.

Joan Bybee. 2010. *Language, Usage and Cognition*. Cambridge University Press, Cambridge, UK.

John Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, UK.

Namhee Lee, et al. 2009. *The Interactional Instinct: The Evolution and Acquisition of Language*. Oxford University Press, Oxford, UK.

Patrick Hanks. 2013. Lexical Analysis: Norms and Exploitations. The MIT Press, Cambridge, US.

R. H. Robins. 1989. *General Linguistics: An Introductory Survey*. Longman, London.

Renqiang Wang and Hemin Chen. 2014. A corpus-based study of the relationship between verbs and constructions: The conventionalization of transitive *sneeze* [J]. *Foreign Language Teaching and Research*, (1): 19-31.

Renqiang Wang and Yu Zhou. 2015. A study of the correlation between heterosemy and frequency in Modern Chinese: A note on the validity of the Principle of Parsimony. *Foreign Language and Literature*, (2): 61-69.

Renqiang Wang. 2006. *An Empirical Study of Word Class Labeling in Chinese-English Dictionaries from the Cognitive Perspective*. Shanghai Translation Publishing House, Shanghai, China.

Renqiang Wang. 2009. Grammatical metaphor and the entry of self-designation senses into Chinese dictionaries: A corpus-based study. *Foreign Language and Literature*, (1): 100-108.

Renqiang Wang. 2010. A validity study of the word class system in Modern Chinese as seen from the

Contemporary Chinese Dictionary (5th edition). *Foreign Language Teaching and Research*, (5): 380-386.

Renqiang Wang. 2013. A study of multiple class membership in Modern Chinese with a comment on the significance of the linguistic theories of Ferdinand de Saussure [J]. *Foreign Language and Literature*, (1): 12-20.

Renqiang Wang. 2014a. Two-level word class categorization in analytic languages: A comparative study of multiple class membership in Modern Chinese and Modern English. In *Proceedings of Workshop of Grammatical categories in macro- and microcomparative linguistics in 36th Annual Conference of the German Linguistic Society*, March 5th-7th 2014, University of Marburg, Germany: 345~347.

Renqiang Wang. 2014b. Multiple class membership in Modern English: A study based on *Oxford Advanced Learner's Dictionary (7th ed.)*. *Journal of Foreign Languages*, (4): 50-59.

Rui Guo. 2002. *A Study of Chinese Word Classes*. The Commercial Press, Beijing, China.

Shiwen Yu, et al. 2003. *The Grammatical knowledge-base of Contemporary Chinese - A Complete Specification*. Tsinghua University, Beijing, China.

William Croft and Eva van Lier. 2012. Language universals without universal categories. *Theoretical Linguistics*, 38(1-2): 57~72.

William Croft. 1991. Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information. The University of Chicago Press, Chicago.

William Croft. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.