

# A Synchronous Corpus-Based Study of Verb-Noun Fluidity in Chinese

Oi Yee Kwong and Benjamin K. Tsou  
Language Information Sciences Research Centre  
City University of Hong Kong  
Tat Chee Avenue, Kowloon, Hong Kong  
{rlolivia, rlbtou}@cityu.edu.hk

## Abstract

The problem of verb-noun categorial ambiguity is critical and relatively unique for non-inflectional languages, especially Chinese. We consider the verb-noun categorial fluidity a continuum and any categorial shift a transitional process. A synchronous corpus-based study was conducted to compare the phenomenon with respect to news texts collected from Hong Kong, Beijing, and Taiwan. It was found that about 15% of the verbs in the Hong Kong and Taiwan texts were undergoing the verb-noun categorial shift; whereas Beijing texts had more than 18% of the verbs undergoing this shift. The results also have important implications on various natural language applications, including lexicography, part-of-speech tagging of Chinese, as well as other natural language processing tasks.

## 1 Introduction

The problem of part-of-speech (POS) or categorial ambiguity, especially between nouns and verbs, as well as adjectives and verbs, has been widely discussed in the literature. Most contemporary Chinese grammarians would suggest that Chinese words have predefined lexical categories mainly based on their syntactic properties, which should not be contingent upon the grammatical function the words take in particular sentences (e.g. Zhu, 2001). For example, the word 懷疑 *huai2yi2*, with a verb sense (“to suspect”) to start with, should only appear as a verb in a dictionary, despite its variable usages found in real contexts, as in (1)<sup>1</sup>.

- (1) a. 我懷疑他是賊 (I suspect he is a thief) -- *verbal, predicate*  
*wo3 huai2yi2 ta1 shi4 zei2*
- b. 他滿臉懷疑表情 (He wears a suspicious look) -- *adjectival, modifier*  
*ta1 man3lian3 huai2yi2 biao3qing2*
- c. 這只是我的懷疑 (This is only my suspicion) -- *nominal, head of object noun phrase*  
*zhe4 zhi3shi4 wo3 de5 huai2yi2*

We call this relative flexibility of a word being used for different grammatical functions and possibly different POSs “*categorial fluidity*”. This phenomenon thus poses a dilemma on tagging Chinese corpora. On the one hand, the theoretical and lexicographic concern is to keep the whole classification of Chinese syntactic categories straightforward, so that all words can be prescribed a category and all categories can be adequately described. On the other hand, the practical concern is to obtain authentic data on the distribution of word uses as this is the ultimate purpose of corpus tagging. Chinese POS tagging in the past had handled this problem in different ways (e.g. CKIP, 1993; Yu *et al.*, 1998; Xia, 2000). For instance, regarding the fluidity between verbs and nouns, Xia (2000) assigned a verb tag if the word is used as a verb and a noun tag if it is used as a noun; whereas CKIP (1993) assigned the same

---

<sup>1</sup> The digits following the Hanyu pinyin indicate the tone.

tag to the same word, but used features to encode specific grammatical functions, such as using [+nom] to indicate a nominalised usage of a verb. We suggest that the two concerns could be compromised, possibly via the tagset, as in Kwong and Tsou (2003), and maintain the mutual relation between the treatment of categorially fluid words and empirical evidence. Thus in the current work we further our empirical study of categorial fluidity between verbs and nouns. Moreover, we try to compare this phenomenon with respect to texts collected from various Chinese speech communities.

In Section 2, we first discuss the problem of POS ambiguity in Chinese. Then in Section 3, we consider the shift of a word from one POS to another a transitional process and identify three types of POS ambiguity along the *categorial fluidity continuum*. In Section 4, we describe the synchronous corpus-based study and report on the results obtained. Finally in Section 5, we discuss the results further with respect to their implications on lexicography, POS tagging, and other natural language processing tasks.

## 2 POS Ambiguity in Chinese

When a word has multiple potential POS tags, the ambiguity has to be resolved in context. The problem of POS ambiguity is especially salient for Chinese, mainly for two reasons. First, categorial change in Chinese words is not often associated with morphological marking. As in (2), the genuine ambiguity between 領導 *ling3dao3* as a verb (“to lead”) and as a noun (“leader”) is not marked by any derivational affixes, unlike its English counterparts.

- (2) a. 領導國家前進 (to lead the country forward) -- verb  
*ling3dao3 guo2jia1 qian2jin4*  
 b. 他是我們的領導 (He is our leader) -- noun  
*ta1 shi4 wo3men2 de5 ling3dao3*

Second, there is no inherent one-to-one relationship between grammatical function and syntactic category. In (3), for example, the word 唱歌 *chang4ge1* (“to sing”) is the predicate in (a) but is the modifier in (b) without any change in its form or indication by function words, unlike (c).

- (3) a. 他唱歌 (He sings) -- predicate  
*ta1 chang4ge1*  
 b. 唱歌技巧 (singing skill) -- modifier  
*chang4ge1 ji4qiao3*  
 c. 唱歌的技巧 (skill of singing) -- modifier  
*chang4ge1 de5 ji4qiao3*

Hence it would not be ideal to classify Chinese words according to their grammatical function in a sentence as far as a classification system is concerned.

The first case above would be relatively easy to handle, as most people would accept it as genuine ambiguity and simply assign the corresponding category according to the sentential context. When 領導 *ling3dao3* refers to the person, it is a noun; and when it refers to the action of leading, it is a verb. The two uses are clearly distinguished. Similar examples include 鎖 *suo3* (noun for the “physical lock” and verb for the “action of locking”), 指揮 *zhi3hui1* (noun for the “conductor” and verb for the “action of conducting”), 關 *guan1* (noun for “a difficult time or challenge” and verb for the “action of closing”), etc.

The second case is more controversial. On the one hand, such variation of usages does not often involve substantial changes in the meaning, and is usually just a difference between the action itself and an abstraction of it. For instance, 領導 *ling3dao3* as in 他領導我們 *ta1 ling3dao3 wo3men2* (he leads us) and 在他的領導下 *zai4 ta1 de5 ling3dao3 xia4* (under his leadership) only differ in the sense that the latter is an abstraction, referring to the activity of leading. On the other hand, it cannot be denied that

in changing the usage, the word has lost some or all of the properties belonging to its original category and acquired some or all of the properties belonging to another category.

There have been a number of different proposals for handling this phenomenon in Chinese. The earliest thoughts were to categorise a word according to its grammatical function in individual sentences, and we have just discussed its impracticality. This would presuppose the absence of inherent categorial property for relevant sentential units. It is also not acceptable as far as a classification system is concerned. Others suggested that the syntactic category of a word should to a certain extent depend on its meaning. If the meaning is the same or similar even if the word is used in different grammatical contexts, its syntactic category should not change. According to this proposal, all the 唱歌 *chang4ge1* in (3) would be tagged as verb. Yet another approach, especially held by contemporary Chinese grammarians, is to determine the syntactic category of a word by the syntactic features it demonstrates or potentially possesses. We will further discuss this in the next section.

Notwithstanding what was said above, according to Liu (2000), POS ambiguity is a serious problem for Chinese but only in a small portion of the Chinese lexicon, especially among the most frequently used lexical items.

### 3 Categorial Fluidity and Shift

Very often when a word is shifted from one lexical category to another, it loses and gains some syntactic features simultaneously. In this study, we will focus on the fluidity between verbs and nouns. For instance, verbs can usually be modified by adverbs and can be followed by tense and aspect markers. In the case of transitive verbs, they can take an object. Nouns, on the other hand, can most prominently be preceded by number and classifier, and modified by adjectives. Thus a nominalised transitive verb cannot take an object in the normal VO order, as in (4).

- (4) a. 影響別人 (to influence others) -- *verb, transitive*  
*ying3xiang3 bie2ren2*  
b. 產生影響 (produce influence) -- *nominalised, object to verb*  
*chan3sheng1 ying3xiang3*  
\*c. 產生影響別人 (produce influence others)  
*chan3sheng1 ying3xiang3 bie2ren2*

At the same time, it gains some syntactic features of nouns, such as modification by adjectives or directly preceded by number and classifier, as in (5)

- (5) a. 一項宣佈 (one announcement)  
*yi1xiang4 xuan1bu4*  
b. 一項重要的宣佈 (one important announcement)  
*yi1xiang4 zhong4yao4 de5 xuan1bu4*

Hu (1996) tested a set of verbs found in dictionaries for verb-noun ambiguity based on three criteria: (i) whether the word can be directly preceded by number and classifier, (ii) whether the word can be directly used as the object of the verb 有 *you3*, and (iii) whether the word can be directly modified by a noun. He found that in a stricter sense, 392 out of 3036 verbs (12.91%) fulfilled at least two of the three test criteria. In a looser sense, however, 587 (i.e. 19.33%) of the test words would be verb-noun ambiguous, fulfilling at least one of the three test criteria.

The asymmetry between nominalisation and verbalisation in Chinese was observed in Tai (1997). In other words, verbs are more freely deverbalised than nouns denominalised. This fluidity between verbal and nominal status of verbs can in theory be generalised to many, if not all, verbs. Nevertheless, does such categorial shift happen to all verbs? What words undergo such shift and how fast? At what point is

the shift significant enough to become genuine ambiguity? In the following we approach these questions with evidence from empirical data.

### 3.1 Three Levels of Ambiguity

To start with, we consider categorial fluidity a continuum, from genuine ambiguities to specific cases of “word play”. A categorial shift is treated as a transitional process, moving from the “word play” end toward the genuine ambiguity end. Thus we distinguish three levels of ambiguity, namely “*regular ambiguity*”, “*transitional ambiguity*” and “*innovative ambiguity*”. By regular ambiguity, we refer to a word having multiple POSs which are well accepted and described in any existing lexicon, as the example of 張 *zhang1* shown in (6) which may be argued to be cases of homonyms or even polysemy. Innovative ambiguity, at the other extreme, accounts for words which are sometimes deliberately used in uncommon ways to create a special effect, as the word 小丑 *xiao3chou3* in (8). Such individual cases cannot be regarded as genuine ambiguity, until the special use becomes common enough. In between the two ends is transitional ambiguity, in which we are most interested for the current study. A word which originally belongs to a particular syntactic category and gradually assumes usage of another category undergoes a process of categorial shift. For instance, many prepositions in Chinese have evolved from verbs, as for the case of 透過 *tou4guo4* in (7).

- (6) a. 一張白紙 (a piece of blank paper)  
*yi1 zhang1 bai2zhi3*  
b. 飯來張口 (rice ready, open mouth)  
*fan4 lai2 zhang1 kou3*  
c. 張先生 (Mr. Zhang)  
*zhang1 xian1sheng1*
- (7) a. 陽光透過窗戶 (sunlight passes through window)  
*yang2guang1 tou4guo4 chuang1hu4*  
b. 透過討論找答案 (through discussion find answer)  
*tou4guo4 tao3lun4 zhao3 da2an4*
- (8) 他很小丑 (He [is] very clown[ish])  
*ta1 hen3 xiao3chou3*

As said, when innovative uses become frequent enough, at some point they might be considered or at least treated as genuine ambiguities. Nevertheless, it is the process of transition that is of linguistic interest, and the usage frequencies from corpus data gathered over time would give good evidence for it. As a result, when tagging a corpus, the different uses of the same word must be sufficiently represented to enable this kind of longitudinal empirical study to be carried out.

### 3.2 Treatment of Nominalised Verbs

For the current study, our working definition of nominalised verbs is corpus-based. A verb is tagged as nominalised if it is found in one of the following five contexts:

- (i) directly preceded by number and classifier,
- (ii) directly modified by adjectival phrase,
- (iii) directly modified by a noun,
- (iv) as modifier in a nominal phrase, or
- (v) directly used as the object of 有 *you3*.

## 4 A Synchronous Corpus-Based Study

### 4.1 Materials

In this study, we used a subset of the LIVAC corpus<sup>2</sup> (Tsou and Lai, 2003). This subcorpus consists of newspaper texts from Hong Kong, Beijing, and Taiwan, covering local news, international news, financial news, sports news, and entertainment news, collected over the *same* period of time. The subcorpus sizes excluding numbers and non-Chinese characters are about 104K word tokens (15K word types) for Hong Kong data, 97K word tokens (13K word types) for Beijing data, and 74K word tokens (12K word types) for Taiwan data. The type-token ratios are similar for all three text sources. A further breakdown by news type is shown in Tables 1-3. The texts were segmented into words and each word was POS-tagged with the approach described in Kwong and Tsou (2003), and the nominalisation of verbs was determined by the criteria listed in Section 3.2.

News Type	Word Tokens	Word Types
Local	15,089	4,297
International	19,670	4,886
Financial	18,898	3,596
Sports	31,518	5,967
Entertainment	18,744	4,517

Table 1 Corpus Size by News Type for HONG KONG DATA

News Type	Word Tokens	Word Types
Local	6,508	1,906
International	17,721	3,477
Financial	16,022	3,390
Sports	36,135	6,172
Entertainment	20,568	4,855

Table 2 Corpus Size by News Type for BEIJING DATA

News Type	Word Tokens	Word Types
Local	8,100	2,498
International	21,503	4,238
Financial	9,858	2,354
Sports	18,509	4,152
Entertainment	15,628	4,010

Table 3 Corpus Size by News Type for TAIWAN DATA

<sup>2</sup> <http://www.livac.org>

## 4.2 Method

First, all verbs (excluding copula verbs and auxiliary verbs) with their frequencies were extracted. Then out of these verbs, those which also exhibited noun or nominalised usages were extracted together with the corresponding frequencies. We call this set of verbs "VN words".

Next, a simple ratio (Eq.1) was computed for all VN words. The *log* ratio was used to give a linear scale. If verb usage outnumbers noun usage to a certain extent, i.e. when  $r \gg 0$ , it suggests that the word is originally a verb and has just started to shift. If verb usage and noun usage are more or less equal, i.e. when  $r \approx 0$ , then either the shift is mature enough or there is genuine ambiguity. Meanwhile, if noun usage outnumbers verb usage by a lot, i.e. when  $r \ll 0$ , it would mean that either the verb has over shifted or the word is originally a noun and is occasionally denominalised.

$$r = \log_2 \frac{\text{verb uses}}{\text{noun uses}} \quad (\text{Eq.1})$$

## 4.3 Results

The general results for the three data sources are shown in Table 4. The percentage of verbs is out of the total number of word types. The column "No. in VN Shift" indicates the amount of VN words out of all verbs. Out of these we analysed those with total frequency (i.e. the sum of both verb and noun usages) 5 or more for their  $r$  values, and the results are shown in the last three columns. It happened that each place had about 60% of the words reaching this threshold. With  $r \geq 1$ , verb usage at least doubles noun usage. With  $1 > r > -1$ , verb and noun usages are quite balanced. With  $r \leq -1$ , noun usage at least doubles verb usage. The general results in Table 4 thus suggest that in real use, there are about 3-4% more nominalised uses of verbs found in Beijing texts than in Hong Kong and Taiwan texts, which indicate quite a different, if not innovative, style of writing in Beijing texts. The figures also reflect the asymmetry between deverbalisation of verbs and denominalisation of nouns.

Source	No. of Verbs	No. in VN Shift	$r \geq 1$	$1 > r > -1$	$r \leq -1$
			<i>(Only for word types with freq<math>\geq</math>5)</i>		
Hong Kong	5,038 (33.8%)	723 (14.4%)	51.6%	26.0%	22.4%
Beijing	3,913 (30.4%)	724 (18.5%)	45.5%	29.7%	24.8%
Taiwan	3,617 (31.4%)	560 (15.5%)	49.1%	27.1%	23.8%

Table 4 Summary of Results (General)

The further analyses for the individual communities with regard to different news types are shown in Tables 5-7 for Hong Kong, Beijing and Taiwan respectively.

News Type	No. of Verbs	No. in VN Shift	$r \geq 1$	$1 > r > -1$	$r \leq -1$
			<i>(Only for word types with freq<math>\geq</math>5)</i>		
Local	1,405 (32.7%)	125 (8.9%)	44.4%	24.1%	31.5%
International	1,647 (33.7%)	148 (9.0%)	51.4%	19.4%	29.2%
Financial	1,083 (30.1%)	142 (13.1%)	45.5%	34.1%	20.5%
Sports	2,121 (35.5%)	225 (10.6%)	52.3%	22.9%	24.8%
Entertainment	1,532 (33.9%)	107 (7.0%)	54.5%	23.6%	21.8%

Table 5 Results Breakdown for HONG KONG DATA

News Type	No. of Verbs	No. in VN Shift	$r \geq 1$	$1 > r > -1$	$r \leq -1$
			<i>(Only for word types with freq<math>\geq</math>5)</i>		
Local	535 (28.1%)	90 (16.8%)	27.1%	33.3%	39.6%
International	984 (28.1%)	151 (13.3%)	40.2%	28.0%	31.7%
Financial	1,036 (30.6%)	156 (15.1%)	35.1%	37.7%	27.3%
Sports	1,925 (31.2%)	235 (12.2%)	36.1%	37.7%	26.2%
Entertainment	1,387 (28.6%)	140 (10.1%)	34.5%	39.7%	25.9%

**Table 6 Results Breakdown for BEIJING DATA**

News Type	No. of Verbs	No. in VN Shift	$r \geq 1$	$1 > r > -1$	$r \leq -1$
			<i>(Only for word types with freq<math>\geq</math>5)</i>		
Local	728 (29.1%)	81 (11.1%)	41.7%	30.6%	27.8%
International	1,333 (31.5%)	169 (12.7%)	42.1%	22.4%	35.5%
Financial	707 (30.0%)	82 (11.6%)	58.5%	17.1%	24.4%
Sports	1,246 (30.0%)	116 (9.3%)	45.8%	32.2%	22.0%
Entertainment	1,286 (32.1%)	110 (8.6%)	49.0%	33.3%	17.6%

**Table 7 Results Breakdown for TAIWAN DATA**

The detailed results in Tables 5-7 echo the observation that Beijing texts demonstrated more nominalised usages of verbs, with an even greater contrast, to the other two places. In addition, we see from the last three columns of the tables that for Hong Kong and Taiwan texts, on average about 50% of the VN words are just beginning to shift (with  $r \geq 1$ ) and their verb usages are still dominant. On the contrary, only about 35% on average of the VN words have  $r \geq 1$  for the Beijing texts. In other words, many words which are verbs originally do not actually play the role of verbs in Beijing texts. This might suggest that Chinese grammar is more seriously Europeanised in the Mainland.

As far as individual text types are concerned, the differences are not particularly obvious, but the phenomenon is apparently very salient in financial news, which ranks first for Hong Kong texts and second for Beijing and Taiwan texts.

As mentioned in Section 3, Hu (1996) tested dictionary entries for verb-noun ambiguity by three criteria and found that 12.91% and 19.33% can have verb usages or noun usages by fulfilling at least two and at least one out of the three criteria respectively. Apparently our findings in this study are similar to Hu's. His data were mainly based on language use in Mainland China, hence our results with respect to the Beijing texts are especially in line with his figures.

## 5 Discussions and Conclusion

The results above suggest that more verb-noun shifts are found from Beijing texts, compared to Hong Kong and Taiwan texts. We further found that within our data, 105 VN words are shared by all three places. There is a moderately strong positive correlation of the  $r$  values, but with a number of lexical items, it is obvious that their usages are quite different among the three places. Some examples and the corresponding  $r$  values are shown in Table 8.

VN Word	<i>r</i> (Hong Kong)	<i>r</i> (Beijing)	<i>r</i> (Taiwan)
發揮	0.8074	2.7004	3.3219
經營	1.5850	-0.1375	1.8074
宣傳	-1.3785	1.5850	0.4854
合作	1.3785	-1.1635	0.4594
衝擊	-0.3219	1.3219	-3.3219
感受	2.8074	-1.4150	-1.3219

Table 8 Examples of VN Words Common to Three Places

Table 8 shows that the word 發揮 *falhui1*, for example, maintains most of its verb usages in Beijing and Taiwan, but is considerably balanced with its noun usages in Hong Kong. On the other hand, the word 感受 *gan3shou4* is mostly used as a verb in Hong Kong texts, but its noun usages predominate in Beijing and Taiwan texts. Some sample sentences from the data are shown in Table 9.

Text Source	Verb Usage	Noun Usage
Hong Kong	(a) 香港單車隊在全運會中發揮合作和犧牲精神。	我清楚他希望踢世界盃，而利物浦的踢法絕對有助他的發揮。
	(b) 難過的時候，她會看看大自然，從一草一木中感受自己的渺小。	Dennis講感受時既感人又爆笑。
Beijing	(a) 聯合國應在這方面發揮重要作用。	申雪由於在賽前熱身活動中扭傷了左踝骨，影響了她在冰面上的發揮。
	(b) 觀眾感受到的是一個群體聲音和形象的轟炸。	這種换位給觀眾帶來的感受很可能是全新的。
Taiwan	(a) 他希望發揮公眾人物的影響力，呼籲大家伸出援手，加入愛心的行列。	……在以防守著稱的意大利隊腳下能有多少發揮空間……
	(b) 演員六月、沈世朋等人與近百位小朋友在寒冬中，提早感受過節的氣氛。	不過，最重要的還是合作的方向和感受。

Table 9 Sample Sentences

In the analysis in Section 4, we did not actually make a distinction for different senses of the same word form and the real nature of the ambiguity. Thus, in Table 10 where we show examples from Hong Kong data under the three bands of *r* values, some of the words might belong to the so-called regular ambiguity, as discussed in Section 3. For example, 省 *sheng3* could mean “province” when used as a noun and “to save” when used as a verb, which are two substantially different meanings. In the future we might need to make such a distinction to better observe the more challenging phenomenon of transitional ambiguity regarding real verb-noun shifts.



Another point which might be of interest is to further analyse the nominalised usages with respect to the corresponding syntactic contexts, such as whether the nominalised verb is used as the head or the modifier in a nominal phrase. For instance, in the current study we did not analyse, for a VO structure, how often it is a verb-object structure and how often it is a modifier-head structure. This is worth further analysis in order to see the actual functions of nominalised usages of verbs.

The results of this study also have important implications on different natural language applications, such as lexicography, POS tagging, as well as other natural language processing tasks. For instance, on lexicography, a threshold could be devised and evidence could be obtained from our empirical data to determine when a verb undergoing the transitional verb-noun shift could be appropriately represented with both POSs in a dictionary. Thus in the last section we have seen that there is a continuum of the verb-noun shift and Table 10 shows examples with different  $r$  values. Hence, words like 意識 *yi4shi2* (“consciousness”), 交易 *jiao1yi4* (“transaction”), and 選舉 *xuan3ju3* (“election”) could possibly be comfortably put under both verbs and nouns given their distribution of actual usages in real texts. For POS tagging, such transitional usages are preferably captured in the tagging process to enable empirical studies like this to be carried out on the one hand, and to distinguish between different syntactic structures such as verb-noun vs modifier-head on the other, which might be an important piece of information for subsequent language processing steps such as parsing.

Moreover, the relation between the extent of verb-noun categorial fluidity and text source as well as text type could be used as a cue for text categorisation. For example, in this study we found that Beijing texts exhibited 3-4% more verb-noun shifts than Hong Kong and Taiwan texts. This could be used as evidence or a feature for recognising writing styles or categorising texts by geographical origins.

$r \geq 1$		$1 > r > -1$		$r \leq -1$	
Word	$r$	Word	$r$	Word	$r$
了解	4.1699	安排	0.8074	意識	-2.0000
愛	4.1699	發揮	0.8074	衝突	-2.0000
懷疑	4.1699	入境	0.8074	聲明	-2.3219
直播	4.0875	支援	0.8074	交易	-2.5850
保留	4.0000	爭議	0.8074	貸款	-2.5850
結束	4.0000	省	0.8074	工作	-2.6114
關注	3.9069	突襲	0.8074	污染	-2.6630
加價	3.8074	保障	0.2226	演說	-2.8074
上升	3.5850	轟炸	0.1926	襲擊	-2.9696
打算	3.5850	上市	0.1769	服務	-3.0000
流通	3.5236	建議	0.1255	運動	-3.0000
出售	3.4594	影響	-0.0317	監製	-3.1699
爭	3.4594	表演	-0.1375	行動	-3.3219
經歷	3.1699	限制	-0.1520	主管	-3.3219
補充	3.1699	規定	-0.1520	指引	-4.0875
希望	3.1197	反擊	-0.2224	選舉	-4.1293
受傷	2.8074	競爭	-0.2224	反應	-4.1699
配合	2.8074	突破	-0.3219	優惠	-4.3923
學習	2.8074	執行	-0.3219	增長	-4.4263
爆冷	2.8074	衝擊	-0.3219	活動	-4.9542

Table 10 Examples of VN Words from HONG KONG DATA

Thus in this paper, we have discussed the problem of verb-noun ambiguity in Chinese based on synchronous corpus data. We treat such categorial fluidity as a continuum and the shift involved as a transitional process. The distribution of such verb-noun shifts was compared across texts from Hong Kong, Beijing, and Taiwan. Results showed considerable variation and indicated that the phenomenon was relatively most common in Beijing text – up to more than 18% of verbs undergo the verb-noun transitional process to various extents. The findings in this study do not only improve our understanding of this perennial problem in contemporary Chinese, but also has important implications for many meaningful natural language applications.

### Acknowledgements

This research is a part of larger efforts of the authors on the study of acquisition of verbs in Modern Standard Chinese by school children in Hong Kong, supported by a Competitive Earmarked Research Grant (CERG) of the Research Grants Council of Hong Kong under grant No. CityU1233/01H.

### References

- Chinese Knowledge Information Processing Group (CKIP). 1993. 中文詞類分析 (三版) Technical Report no.93-05, Academia Sinica, Taiwan.
- Hu, M. (胡明揚) 1996. *Cilei Wenti Kaocha* (詞類問題考察). 北京語言學院出版社.
- Kwong, O.Y. and Tsou, B.K. 2003. Categorial Fluidity in Chinese and its Implications for Part-of-speech Tagging. In *Proceedings of the Research Note Session of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary, pages 115-118.
- Liu, K. (劉開瑛) 2000. *Zhongwen Wenben Zidong Fenci He Biaozhu* (中文文本自動分詞和標注). Beijing: Commercial Press.
- Tai, J.H-Y. 1997. Category Shifts and Word-Formation Redundancy Rules in Chinese. 《中國境內語言暨語言學》, 3: 435-468.
- Tsou, B.K. and Lai, T.B.Y. 2003. Hanyu Gongshi Yuliaoku Yu Xinxu Kaifa (漢語共時語料庫與信息開發). To appear in 《漢語自然語言處理若干重要問題研討》, 國家“973”項目專家組, 北京.
- Xia, F. 2000. *The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)*. IRCS Technical Report IRCS-00-07. Institute for Research in Cognitive Science, University of Pennsylvania.
- Yu, S., Zhu, X. and Li, F. 1998. Representation of Grammatical Knowledge in the Chinese Lexicon. In B.K. T'sou, T.B.Y. Lai, S.W.K. Chan and W.S-Y. Wang (Eds.), *Quantitative and Computational Studies on the Chinese Language*, pages 353-372.
- Zhu, D. (朱德熙) 2001. *Xiandai Hanyu Yufa Yanjiu* (現代漢語語法研究). Beijing: Commercial Press.