From etymology to modern phraseology:
A corpus-based study of structural variants of
Chinese idioms in naturally-occurring contexts

Meng Ji

Department of Humanities, Imperial College London,

South Kensington Campus, London, W5 4RB, UK

m.ji@imperial.ac.uk

Abstract

Compared with recent developments in English corpus lexicology and phraseology, the study of Chinese lexicology and phraseology still remains at a level similar to what Fernando has described as quasi-lexicography (Fernando, 1996: 10-11) in the study of English idioms, where explanations provided in idiom dictionaries are rather prescriptive and static than descriptive and dynamic: a typical template of explanation consists in no more than a paraphrase meaning; the etymological origin of the entry; one or two illustrative examples and synonyms equally in a four-character idiomatic pattern if available. As a result, items collected in recent Chinese idiom dictionaries are presented as if they had not experienced morpho-syntactic variations or pragmatic shifts of any sort since they first came into use hundreds of years ago. The current situation of Chinese phraseology has largely restricted our understanding of the nature of Chinese idioms to a preliminary level of analysis as characterized by subjective or intuition-oriented judgements. Therefore, in the present study, I shall propose for the first time to situate the study on an essential subject of Chinese phraseology, namely the structural variability of Chinese idioms, within a naturally-occurring context as provided by the Corpus of Chinese (Beijing University, 2005), with a view to submitting the detected structural variants of Chinese idioms to a systematic description and to examine whether the seemingly idiosyncratic behaviour of Chinese idioms as highly responsive to the changing contextual circumstances is in fact governed by a set of underlying linguistic rules operating simultaneously at syntactic, semantic and pragmatic levels (Minsky, 1980: I; Fillmore, 1976: 25; Brown & Yule, 1983: 236; Moon, 1998: 163-166). The results obtained in my investigation on the structural versatility of certain kinds of Chinese idioms are in the nature of a pilot study; and the corpus-driven approach adopted in describing and analysing the structural features of Chinese idiomatic variants in naturally-occurring contexts will help prepare the ground for the setting-up of an appropriate methodological framework for future research on the subject matter.

Keywords

1. Introduction

In recent times, a number of large-scale balanced or domain-specific Chinese corpora have been set up and developed into sophisticated text-mining systems in major research centres or educational institutions in and out of the country, including the *Sinica Corpus of Chinese* by the Institute of Linguistics at the National Academy of Taiwan (1996) as representative of the corpus research of traditional Chinese; the ever-expanding *Penn Chinese Treebank* by Pennsylvania University (2000); the well-framed *Lancaster Corpus of Mandarin Chinese* by the Corpus Linguistics Centre at Lancaster University (2004); and the *Corpus of Chinese*, the most recent and largest corpora of modern mandarin Chinese launched into public domain by Beijing University in 2005, to name but a few. The remarkable development gained in computational linguistics with regard to Chinese natural language processing has provided important technical support for the construction and versatilization of Chinese corpus material, signifying the advent of modern Chinese linguistics as characterized by the essential presence of computational or corpus linguistic tools in the description and analysis of the language.

   To a greater extent, corpus-based linguistic studies differ from traditional intuition-oriented research traditions. That is because, in doing corpus linguistics, massive crude though objective natural language data will substitute the subjective citation or enumeration of relevant linguistic evidence by individual linguists in preparing the ground for any analytical argument to be made on a certain linguistic topic. Through the observation of how linguistic entities behave in naturally-occurring contexts, a great deal of novel and revealing information gradually comes to the surface proving important linguistic insights for a deeper and more comprehensive understanding of the language, and from time to time, questioning the validity or suitability of traditional linguistic theories which were used to be generally endorsed.

   In the last three decades or so, as promoted by computer-assisted corpus tools, English linguists have achieved substantial fruits in the study of the so-called *lexical phenomena*, as opposite to syntactic phenomena which are supposed to be largely explainable within the framework of generative grammars; and being an integral part of general lexical study, modern English phraseology has been informed significantly by corpus-based research schema. For instance, based on new facts bearing on the morpho-syntactic structural viability of classified English idiomatic expressions as emerged from corpus database, a set of important theoretical propositions has been formulated and advanced to help explain the rationale behind such complex linguistic phenomena (Fillmore, 1988 & 1994; Nunberg, 1994; Biber, 1993 & 1998; Moon, 1998 & 2000; Hanks, 2003; Sinclair, 2003, etc.), which has led to a radical conceptual shift in the way we perceive of the notion of idiomaticity as embodied in the various types of phraseological units; and theoretical insights gained in this respect have been soon applied by lexicographers into the practice of dictionary design and compilation (Fernando, 1996: 9-13; Hanks, 2003: 48-70).

   Compared with recent developments in English corpus lexicography and phraseology as outlined above, the study of Chinese lexicography and phraseology still remains at a level similar to what Fernando has described as *quasi-lexicography* (Fernando, 1996: 10-11), where explanations provided in Chinese idiom dictionaries are rather prescriptive and static than descriptive and dynamic: a typical template of explanation consists in no more than a paraphrase meaning; the etymological origin of the entry; one or two illustrative examples and synonyms equally in a four-character idiomatic pattern if available. As a result, items collected in recent Chinese idiom dictionaries are presented as if they had not experienced

morpho-syntactic variations or pragmatic shifts of any sort since they first came into use hundreds of years ago; and such state-of-the-art in Chinese phraseology has largely restricted our understanding of the nature of Chinese idioms, which are subject to diachronic developments and synchronic variations just like any other languages that I would suggest. Therefore, in the current study, I shall propose for the first time to situate the study on an essential subject of Chinese phraseology, namely the structural variability of Chinese idioms, within a naturally-occurring context as provided by the *Corpus of Chinese* (Beijing University, 2005), with a view to submitting the detected various types of structural variants of Chinese idioms to a systematic description and to examine whether the seemingly idiosyncratic behaviour of Chinese idioms as highly responsive to the changing contextual circumstances is in fact governed by a set of underlying linguistic rules operating simultaneously at syntactic, semantic and pragmatic levels (Minsky, 1980: I; Fillmore, 1976: 25; Brown & Yule, 1983: 236; Moon, 1998: 163-166).

2. Underlying morpho-syntactic structures of Chinese idioms (*Cheng Yu*)

As Moon suggests that finding exploitations and variations of fixed expressions is the hardest part of corpus-based investigations; and that searches are most successful when the query consists of two lexical words, fairly close to each other (Moon, 1998: 51). In more sophisticated cases of the flexible use of idiomatic expressions, such as truncations, ellipsis, idiosyncratic permutations, despite the various querying methods deployed, important however sparsely distributed data cross large-scale corpora may still easily escape from one's searching scope. Given the state of the art of lexical acquisition in Chinese natural language processing, and the substantial technical limitations implied in a corpus-based approach to the identification of all sorts of variation possibilities that Chinese idioms may display in actual communicative contexts, it should be noted that the results obtained in my investigation on the morphosyntactic versatility of certain kinds of Chinese idioms are in the nature of a pilot study, and the largely descriptive approach adopted in analysing the constructional peculiarities of Chinese idiomatic variants hopefully will help prepare the ground for the setting-up of a proper methodological framework that could be used in future research on the topic.

   A remarkable morpho-syntactic feature that most Chinese idioms (or *Cheng Yu* as we say in Chinese) seem to bear is that compared with roughly equivalent phraseological concepts in other languages like *idioms* in English or *phrases hechas* in Spanish, the syntagmatic configuration of *Cheng Yu* is largely restricted to a four-morpheme-length rule, rather than being freely composed. Here, by using the term *morpheme* instead of *word* or *character* to describe the orthographic representation of *Cheng Yu*, I am hoping to avoid the potential confusion that an interchange of these three terms may cause, since so far as I am aware, each of them seems to refer to a different set of linguistic entities with sometimes overlapping though not entirely identical syntactic functions within a clausal or sentential context. Therefore, in the current study, I shall use the term *character* to describe a recorded lexical item in modern Chinese, which may well be single, bi-, tri- or tetra-morphemic long (Chinese characters made up of more than four morphemes do exist, such as transliterations of foreign words, but they are supposed to represent only a very tiny proportion in modern Chinese). Here, as we can see, the notion of *character* exploited in my study resembles to that of *word* in most alphabetical linguistic systems. Lastly, the concept of *morpheme* shall be reserved for an orthographic symbol with basic semantic meaning, as we say *zi* in Chinese.

   In modern Chinese, the frequency of occurrence of monosyllabic morphemes to be used as independent semantic units in textual or discursive contexts is much lower than that of

disyllabic or trisyllabic characters. Such linguistic phenomenon has proved to be in striking contrast with ancient Chinese, where the presence of monosyllabic words was predominant, and serves as important evidence of how Chinese lexicon evolves during centuries from semantically highly synthetic towards increasingly descriptive and analytical through a range of morphological variations of existing monosyllabic words, such as repetition, juxtaposition, modification, etc. (Li, 2002: 68-76). The phraseological type of *Cheng Yu*, which may be seen as a perfect reflection of highly compact morphosyntactic structures of formal ancient Chinese, tends to involve a substantial use of monosyllabic morphemes rather than disyllabic words, given the limited space provided by the largely institutionalized four-character morphosyntactic patterns. In this sense, it could be said that the notion of *morpheme* is much more effective than *character* in describing the morphosyntactic configuration of *Cheng Yu*.

It seems that little in-depth research has been carried out on the etymological origin of the generalized four-character pattern of *Cheng Yu*, which represents some ninety per cent of Chinese idioms (Xu, 2006: 108); however, it is commonly held among Chinese linguists that this unique four-character pattern of Chinese phraseology is a perfect manifestation of the prosodic musicality, structural equilibrium or symmetry, and semantic succinctness of Chinese (ibid, 107). Although theoretically speaking, the unique morpho-syntactic structure of Chinese idioms in terms of the fixed length of each item per se does not have a serious impact on the possibilities for the coinage of new idiomatic variants in a naturally-occurring context, it should be noted that the kinds of structural variations as have been observed and discussed extensively with regard to the flexible use of English idioms, such as modification, quantification, topicalization, permutation, and other grammatically-motivated changing modes like passivation, inflectional variation, etc. (Nunberg, 1994; Fernando, 1996; Moon, 1998; Riehemann, 2001) do not seem to lend themselves so well to the flexible use of Chinese idioms.

The rationale behind this can be tentatively suggested as two-fold: in the first place, Chinese is a typical non-inflectional language, as a result, many kinds of structural variations as derived from grammatical changes are virtually incompatible with Chinese idioms; secondly, as mentioned above, the four-character-long pattern of wording is believed to carry a maximum aesthetic and rhetoric value in Chinese texts, therefore, any sort of structural modification that may go beyond or break down this generalized fixed pattern is definitely less preferred in the creation of Chinese idiomatic variants. In other words, with an English idiomatic variant, one may well expect the underlying pattern of an idiomatic expression to be modified, permuted, truncated, distributed over several clauses, or even totally subsumed into the context, radical structural alterations of these sorts appear to be less common in the flexible use of *Cheng Yu*, for these types of distanced structural variations implicitly go against the institutionalized use of Chinese idioms within four-character patterns.

However, this is not to say that relatively freely composed variants of *Cheng Yu* are totally absent in the real use of the language; by contrast, a pragmatically-motivated flexible use of a Chinese idiom in a larger clausal or sentential context may sound absolutely natural under certain circumstances and serves as a good reflection of the linguistic competence of a native speaker (Ji, 2007 forthcoming). Part of the reasons for leaving these rather elusive freely-structured idiomatic variants out of the scope of the current study has been due to considerations on both the technical operationality in retrieving them from large-scale corpora and their less representative statistical significance as compared with idiomatic variants built on productive schematic constructions. That is, most morphosyntactic variations of Chinese idioms normally take place within the underlined four-character patterns. This specific type of Chinese idiomatic variants is somehow analogous to what

Fillmore (1988) has designated as *formal idioms or grammatical constructions* (1988: 506), which allow parts of the idiomatic content to be instantiated by additional semantic elements while retaining the defining structure of the idiom intact, such as XAXB, AXBX, XXAB, AXBY or AXAY, to name but a few. Thus, they can hardly be seen as completely lexically specific.

   Regarding the selection of a handful of appropriate morphemes from a set of seemingly endless lexical possibilities to fill in the left out positions in an underlying Chinese idiomatic construction, there is still an element of conventionality involved, which despite its extreme importance for understanding the machinery for generating Chinese idiomatic variants in naturally occurring contexts, has rarely been addressed in the literature. The conventionality governing the selection of appropriate morphemes for the creation of new idiomatic expressions actually touches upon a very complex topic. In her study on English fixed expressions (FEIs), Moon proposes to treat the entrenched network of important linguistic constraints imposed on the productive process of idiomatic variants as *idiom schema* (Moon, 1998: 161), which is intended as an integrated approach to such complex phenomena and is claimed to be in line with Fillmore's *semantic frames* (1976:25) and Minsky's *frame theory* (1980:1-25). Within the tradition of frame theory, the proliferation of idiomatic variants derived from a basic structural pattern is treated as subject to a set of correlated rules operating simultaneously at semantic, syntactic and pragmatic levels; and a central task of frame-theory-oriented research is to identify the prosodic and syntactic characteristics and the particular semantics and pragmatics that may differentiate a certain idiomatic structural pattern or construction from others in its class (Fillmore, 1988: 516). It seems that such research schema seeking to adopt a holistic approach to the lexico-syntactic configuration of productive idiomatic patterns has been rarely applied to the study of Chinese idioms; however, it is my contention that in an effort to get a closer and deeper understanding of the nature of Chinese idioms, it is deemed to be necessary to observe and analyse the apparently idiosyncratic behaviour of Chinese idiomatic variants in the real use of the language along the dimensions as proposed by frame theory.

3. A corpus-based approach to Chinese idiomatic variants

To attest the productivity of potentially prolific Chinese idiomatic structures, I am using the *Corpus of Chinese* (CMC), which has been developed by the Centre of Chinese Linguistics at Beijing University[1] and launched into the public domain in January 2005. It is the largest monolingual corpus of modern Chinese ever constructed in mainland China, continuously expanding with the addition of corpus materials from a wide range of sources[2]. The CMC is totally free for online concordance searching and basic information of the CMC regarding the distribution of corpus texts as well as their sources is shown in the following diagram: the main categories of corpus texts compiled in CMC are (counting from left to right along the horizontal axis) *People's Daily* (the official organ of P.R.C); *Digest of Literary Writings*; *Market News and Reports*; *Readers* (magazine); Spoken Chinese; Practical Writing in

---

[1]   The   Corpus   of   Chinese   is   available   at http://ccl.pku.edu.cn:8080/ccl_corpus/jsearch/index.jsp?dir=xiandai, my last access to the CMC was in October 2006

[2] The current size of CMC is of some 229, 700, 435 bytes (219 MB) corresponding to around 115 million Chinese words.

Chinese; Literature (I); Translations; Dictionaries; Drama; Literature (II), where the first nine categories are in contemporary Chinese and the last two in modern Chinese[3].

Diagram I Distribution I of materials in Corpus of Centre of Chinese Linguistics (PKU)[4]

北京大学汉语语言学研究中心现代汉语语料：共有 157 个文件；23 个文件夹；229,700,435 字节（219MB）。约 1.15 亿字。
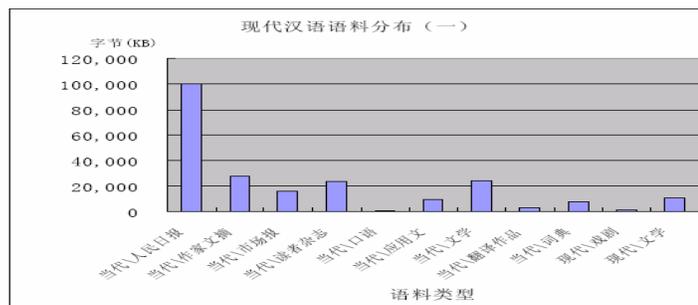
1. VN₁ + VN₂ – Double VN Phrasal Structure

The underlying pattern of many Chinese idioms can be expressed as $VN_1 + VN_2$ for two highly-compacted juxtapositional verbal phrases, with each half containing a monosyllabic verbal morpheme and a monosyllabic noun morpheme. An interesting point worth noting here is that some Chinese idioms are characterized by the use of morphemes in ancient Chinese with shifting grammatical functionalities known as 词类活用, e.g. nouns or adjectives used as verbs; adjectives used as nouns or adverbs; or verbs used as nouns. Therefore, as a result, in the idiomatic construction of $VN_1 + VN_2$, V (verb) and N (noun) should be understood as the grammatical function of the morpheme when it is used in an idiomatic context. For instance, in the Chinese idiom 同甘共苦 (tóng-gān-gòng-kǔ), the $VN_1 + VN_2$ structure is realized through the conversion of the grammatical function of the second and fourth morphemes from adjectives into nouns.

   Drawing on the semantic relations held between the two verbal phrases VN_1 and VN_2, two major subclasses have been established which are (1) VN_1 and VN_2 are semantically synonymous or relevant; and (2) VN_1 and VN_2 are semantically opposite. Under such a taxonomic scheme, I proceed to check on the productivity or versatility of each Chinese idiomatic construction (CIC) falling under the double VN phrasal structure, mainly relying on data retrieved from the *Corpus of Modern Chinese* (CMC). Instances extracted from the CMC sharing a certain underlying CIC usually involve two major kinds of idiomatic expressions, which are conventionalized idioms and idiosyncratic variants of the former. Although with the help of Chinese idiom dictionaries, one may reasonably tell them apart, this is not supposed to be necessary in the present study, which is to attest the productivity of

---

[3] It is generally accepted in China today that the concept of modern Chinese refers to the Chinese used ever since the May Fourth Movement in 1919, while the notion of contemporary Chinese was established in 1949 when the P.R.C was founded.

[4] See the main webpage of CCL of Beijing University at the same WWW address given above

a certain CIC by gathering as many idiomatic items as one possible can from large-scale corpora, regardless of whether they are conventionalized expressions or not, since both established idioms or newly coined idiomatic variants can serve as ample evidence for the structural variability of a given CIC.

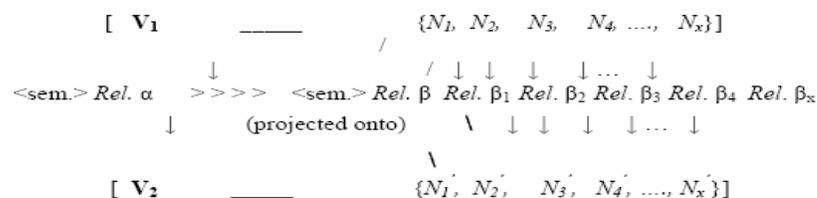## 1.1. VN$_1$ and VN$_2$ are semantically synonymous or relevant

● **{改 (change)…换 (change)…}**

1.1 改[头]换[面] (gǎi-tóu-huàn-miàn) (74)[5]; 1.2 改[天]换[地] (gǎi-tiān-huàn-dì) (37); 1.3 改[朝]换[代] (gǎi-cháo-huàn-dài) (66); 1.4 改[名]换[姓] (gǎi-míng-huàn-xìng) (44); 1.5 改[姓]换[名] (gǎi-xìng-huàn-míng) (1); 1.6 改[颜]换[貌] (gǎi-yán-huàn-mào) (1); 1.7 改[辙]换[戏] (gǎi-zhé-huàn-xì) (1); 1.8 改[妆]换[容] (gǎi-zhuāng-huàn-róng) (1); 1.9 改[祖]换[宗] (gǎi-zǔ-huàn-zōng) (1);

The underlying pattern of this CIC is characterized by the juxtaposing of two semantically identical verb morphemes 改 and 换, which in modern Chinese are sometimes used together as a single disyllabic verb holding the same meaning as *change* or *alter*. The semantically corresponding relation between the two verb morphemes has an important implication on the two positions left out for the underspecified noun morphemes to be filled in the construction. As shown in the ten instances listed above, the semantic meaning of two noun morphemes is invariably analogous to each other, like in 1.1 头 (head) and 面 (face); 1.2 天 (heaven; sky) and 地 (earth); 1.3 朝 (dynasty) and 代 (dynasty); 1.4 名 (given name) and 姓 (surname); 1.5 型 (type) and 代 (generation, both used specially for machine production); 1.6 姓 (surname) and 名 (given name); 1.7 颜 (countenance) and 貌 (face); 1.8 辄 (quantifier for a traditional Chinese play) and 戏 (play or show); 1.9 装 (clothe; dress) and 容 (looks); and in 2.0 祖 (ancestor) and 宗 (ancestor). That is, the semantic relation between the two underspecified noun morphemes N$_1$ and N$_2$ should be seen as largely derivational from that of the two underlined verbs of this CIC, which is schematized in the diagram show below (Diagram II). As a result, the two juxtaposed clauses VN$_1$ and VN$_2$ assert largely identical propositions from two slightly different points of view, one of which reinforces the other, both together constituting a combined assertion regarding a particular social phenomenon (1.3) or agent action (1.1; 1.4; 1.6; 1.7; 1.8; 1.9 and 1.10).

---

[5] The number given in brackets represents the frequency of occurrence of that item in the CMC.

Diagram II Semantic relations between constituent parts of the CIC {改…换…}

$$[\quad V_1 \quad \underline{\quad\quad} \quad\quad\quad\quad \{N_1, \ N_2, \quad N_3, \quad N_4, \ ...., \ N_x\}]$$

$$\downarrow \qquad\qquad\qquad\qquad / \ \downarrow \ \downarrow \quad \downarrow \quad \downarrow \ ... \quad \downarrow$$

$$<sem.> \ Rel. \ \alpha \quad >>>> \quad <sem.> \ Rel. \ \beta \ \ Rel. \ \beta_1 \ Rel. \ \beta_2 \ Rel. \ \beta_3 \ Rel. \ \beta_4 \ Rel. \ \beta_x$$

$$\downarrow \qquad (projected \ onto) \qquad \backslash \ \downarrow \ \downarrow \quad \downarrow \quad \downarrow \ ... \quad \downarrow$$

$$[\quad V_2 \quad \underline{\quad\quad} \quad\quad\quad\quad \{N_1{}', \ N_2{}', \quad N_3{}', \quad N_4{}', \ ...., \ N_x{}'\}]$$

Now, let us have a closer look at the collocational features of this construction {改…换…} to see whether its flexible use in actual communicative contexts is governed by any pragmatic rules. The idiom selected as the starting point is 改头换面 (1.1), being the most commonly used among the ten idioms or idiomatic variants detected from the CMC. Relying on the massive collocation samples extracted from the CMC, we can see that the seemingly neutral use of the idiom 改头换面 turns out to be mostly deployed in a disapproving textual context, which may sometimes be referred to in an ironic tone. The following twenty concordance lines are selected at random from the CMC which show us some revealing findings regarding the pragmatic use of this idiom in naturally-occurring contexts:

Concordance I 改头换面

1. 将"台独"分子提出的"台湾命运共同体"	改头换面	为鼓吹在岛内建立"生命共同体"； 将民进党提出的"国民党政权是外来政权
2. 在各类报刊上开数以十计的言之无物的专栏，或者把若干文章	改头换面	一再编选，重复出版；有的高价走穴的歌星， 竟然也可以堂而皇之地以录
3. 中国反对任何形式的霸权主义和强权政治；反对	改头换面	的新殖民主义和为老殖民主义张目； 反对外来势力干涉非洲国家内部
4. 而且是你中有我，我中有你，油兑水，水掺油；或者是	改头换面	、包装包装，金玉其外。
5. 对于民族虚无主义的全面的清算以及对于８０年代中期这一思潮	改头换面	的出现所做的批判， 有力地推动了"古为今用"方针的执行和贯彻。
6. 仍是魁北克省的执政党，一有时机，魁北克独立势力还可能	改头换面	提出新的全民公决。
7. 也不寄样书或付稿费，有的将他人作品	改头换面	出版或肆意剽窃、抄袭， 严重侵犯著作者及版权所有单位权益。
8. 中央制定的有关政策难以落实，许多优惠政策被地方随意	改头换面	挪用，农民得不到实惠； 四是基层有关部门为农民服务差，有的还乱收费。
9. 一名公安战士身负重伤昏睡多年后苏醒，得知当年的罪犯	改头换面	，仍在危害社会，便毅然重入缉拿凶犯险途的故事。
10. 挂着『人体科学』招牌，行算命迷信之实，现在又出	改头换面	招摇过市。

现老黄历		
11. 旬邑苹果却被一些商贩收去充作"商品表皮"、"搭面子"、"	改头换面	"、"改名换姓"。
12. 某处被禁止入网的信息，可易地进入网络，或	改头换面	以另一种形式进入网络。
13. 商业广告正以锐不可当之势围剿和肢解着成语，那些被	改头换面	，被抽筋断骨，带有浓厚商业色彩的畸形成语，堂而皇之、有声有色地走进了
14. 也有的把外国的一些不入流的宗教方面的图书，	改头换面	，标榜谈哲理、讲人生，貌似深刻其实是些浅显平庸的读物。
15. 因为我们党内有10次路线错误，今后还会有的，敌人会	改头换面	藏在我们的党内"。
16. 以"参考"为名，将先出的成名译本拿来，正改几句，歪改几行，	改头换面	，滥充新译。
17. 许多材料不是深入进行实际调查研究所获，而是靠剽窃、靠	改头换面	摘抄他人现成资料而为。
18. 这是经济利益驱动下的变相评奖，是不是前些年搞评奖的人	改头换面	重操旧业？
19. 过去的厂长、经理	改头换面	，又成了董事长兼总经理，依旧大权独揽。
20.查禁的宣扬色情、凶杀暴力、封建迷信和反动内容的出版物，经	改头换面	，重新出笼；有的已被撤销的出版单位还在进行违法犯罪活动

From the top line to the bottom line, the agents being described by the idiom 改头换面 serving as a situational modifier are separatists of Tai Wan (1.); meaningless publications (2.); hegemony or imposing politics (3.); fake petrol products (4.); national nihilism (5.); separatists of Quebec (6.); plagiarized works (7.); preferential policies of central government to farmers abused by corrupt local authorities (8.); criminals (9.); superstitious beliefs (10.); food supply of bad quality but in decent decorations (11.); prohibited web information (12.); misuse of proper Chinese in commercial advertisements (13.); fake religious reading material (14.); political opponents (15.); plagiarized translations (16.); fake investigation results (17.); people organizing appraisal activities as purely motivated by commercial interests (18.); monopolizing administrative staff (19.); and pornographic, violent or superstitious print material (20.). As can be seen, in some fourteen cases out of the twenty samples (2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 16, 17, 20), the agent of action being modified by this idiom is non-personal or non-animate, thus the idiom is used in its metaphorical sense which leads to an ironic interpretation of its literal meaning in the textual context.

These corpus data seem to suggest that in devising an idiomatic variant on the construction of {改…换…}, one would at least need to bear in mind two basic linguistic factors restricting the flexible use of this lexically semi-filled pattern, namely the semantic symmetry of the two underspecified noun morphemes and the somehow conventionalized pragmatic use of this construction as a disapproving and ironic phraseological pattern in many communicative contexts. In attempt to corroborate such an assumption bearing on the two highlighted dimensions of idiomaticity of this construction, i.e. semantic and pragmatic, let us have a second look at the collocational features of the five idiomatic variants (1.5 – 1. 9) detected from the CMC, which only seems to confirm such a hypothesis of mine.

Concordance II Other idiomatic expressions constructed on the CIC {改...换...}

1.5 非得常常	改姓换名 不可。
1.6 随时可能被红卫兵揪住辱骂的处境，变为了具有每到一处，便能使该处事前	改颜换貌 的法力。
1.7 如果发现哭声和眼泪已引起对方不耐烦、反感，她能马上刹车，	改辙换戏
1.8 者俯瞰街市，便会发现朝朝暮暮沸腾生活的流动和都市日日夜夜在	改妆换容
1.9 封建思想'和'极左'观点，我宁肯永远背着这两大'罪名'也不想	改祖换宗

Although very limited information is provided in the short concordance line 1.5 without any reference to the subject of the action indicated by 改姓换名 (1.5), which is a permutation variant of the idiom 改名换姓 (1.4) (change-given name-change-surname), the ironic effect brought about by this CIC variant in the context can still be hinted at the use of the adverb 常常 (often) to modify the action denoted by改姓换名 (change-surname-change-given name). Similarly, with the exception of the sample 1.8, the contextual information given in the concordance lines 1.6, 1.7 and 1.9 serves as very illustrative evidence for the necessity of an ironic or derogatory reading of the use of idiomatic variants built on this CIC in actual context. For examples, in the sample 1.7, the use of the expression 改辙换戏 (change-play-change-show) is intended as a mocking reference made to the behavior of a cunning woman; in the sample 1.9, 改宗换祖 (change-ancestor-change-ancestor) may be construed as a metonymic reference to the conversion of one's original ideological or political stance, as a result of which the ironical tone of the expression is implicit.

In the sample 1.6, 改颜换貌 (changed-countenance-changed-face) is used as a modifying element to describe the object noun 法力 (magic capacity), a surviving skill acquired by the subject after being persecuted on several occasions by red rebels in the Great Cultural Revolution. Here the pragmatic connotation of the expression 改颜换貌 as ironic or humorous, rather than being context-free, largely depends on the availability of the collocate法力 (magic capacity) in the same textual context. Somehow different, the use of a similar expression 改装换容 (1.8) (change-cloth-change-looks) to改颜换貌 (changed-countenance-changed-face) in the sample sentence 1.8 proves to be in a rather neutral or even appraising way: it seems to have been used in a newspaper article to report the ever-changing appearance of cities in a fast-growing economy; and in such a contextual environment, the explanation of a certain constituent word should only be approached as positive, given that the underlying tone of the whole text is affirmative. The contrast reflected in the comparison of the pragmatic uses of two items built on the same construction seems to suggest that although conventionality plays a vital role in the flexible use and interpretation of Chinese idiomatic variants, a proper understanding of any word in real language use should never be attempted without making reference to the textual context.

1.2 VN$_1$ and VN$_2$ are semantically contrastive or opposite

- {有 (with)… 不/无 (without; no)…}

2.1 有[条]不[紊] (yǒu-tiáo-bù-wěn) (193); 2.2 有[气]无[力] (yǒu-qì-wú-lì) (124); 2.3 有[恃]无[恐] (yǒu-shì-wú-kǒng) (71); 2.4 有[惊]无[险] (yǒu-jīng-wú-xiǎn) (50); 2.5 有[名]无[实]

(yǒu-míng-wú-shí) (50); 2.6 有[备]无[患] (yǒu-bèi-wú-huàn) (34); 2.7有[始]无[终] (yǒu-shǐ-wú-zhōng) (15); 2.8 有[眼]无[珠] (yǒu-yǎn-wú-zhū) (12); 2.9 有[教]无[类] (yǒu-jiào-wú-lèi) (7); 2.10有[勇]无[谋] (yǒu-yǒng-wú-móu) (10); 2.11有[头]无[尾] (yǒu-tóu-wú-wěi) (7); 2.12有[才]无[德] (yǒu-cái-wú-dé) (5);

This CIC is basically characterized by the semantically opposite relation holding between the verb morphemes as 有 (have; with) and 不/无 (no; not; without). A first look at the noun morpheme pairs filled in this lexically semi-specified morpho-syntactic pattern shows that as in the case of {改(change)…换(change)…}, here the semantic oppositeness existing between the two fixed verb morpheme has largely prescribed the nature of lexicons to substantiate this CIC; that is, the two nouns should equally be semantically contrastive or opposite:

条 (order) and 紊 (disorder) (2.1); 气 (breath) and力 (strength) (2.2); 恃 (rely/depend on) and 恐 (fear; dread; restraint) (2.3); 惊 (panic; alarm) and 险 (danger) (2.4); 名 (name; prestige) and实 (real/ actual strength) (2.5); 备 (preparation) and 患 (worry) (2.6); 始 (start) and 终 (end) (2.7); 眼 (eye) and 珠 (eyeballs; discerning capacity) (2.8); 教 (education) and 类 (types) (2.9); 勇 (courage; daring) and 谋 (intelligence) (2.10); 头 (start) and 尾 (end) (2.11); 才 (knowledge) and 德 (virtues) (2.12).

However, if we look more closely at these combinations, we would discover that the generally described semantic contrastiveness between these paired noun morphemes cannot be taken in a straightforward manner, for the pragmatic importance conveyed by each part is imbalanced rather than parallel. To be more precise, the semantic frame of this CIC implicitly gives more emphasis on the significance of the quality represented by the second noun morpheme in making an evaluative assessment of an individual or a certain situation; and it is suggested that in grasping the real meaning of idioms or idiomatic expressions constructed on this morphosyntactic mould, one should focus on the second half of this semi-specified pattern as {不/无 (no; without)…}, rather than on the first half as {有 (with)….}, which in many cases is limited to provide a second-class contrast to make more salient the quality indicated by the second filled-in morpheme, since the latter tends to be more valued or appreciated by the speaker.

  For instance, in the following eight samples 有恃无恐 (2.3), 有惊无险 (2.4), 有名无实 (2.5), 有始无终 (2.7), 有眼无珠 (2.8), 有勇无谋 (2.10), 有头无尾 (2.11), 有才无德 (2.12), it is intuitively obvious from my point of view as a native speaker that the intended purpose of the speaker in arranging the positions of the two noun morphemes in each item as such is to transmit a clear message to his/her interlocutor "look, although he/she seems to have the quality of A (represented by the first noun morpheme), it is a shame that he/she does not possess the quality of B (represented by the second noun morpheme), which I believe is more important than A". Similarly, if the expression is used regarding the assessment of a situation like in 2.4, 2.7 and 2.11, the physical priority enjoyed by the first noun morpheme in the construction of {有…不/无…} may well have a reverse impact on its semantic significance in actual communicative context. For instance, in有惊-无险 (2.4), it is implicit in the speaker's intention to say that despite the situation being fragile or full of potentially uncertain factors, it has turned out to be in control without real threats or dangers, thus the emphasis is put on 无险 (without danger) instead of 有惊 (with panic; alarm).

  Based on observations of the pragmatic properties of the CIC {有…无…}, we could probably argue that the two constituent parts VN_1 and VN_2 are prosodically coordinated with each part containing a focused element, N_1 and N_2. Within this paired focus construction, the

semantic relation between N_1 and N_2 is imbalanced rather than merely contrasting; and such semantic imbalance is reflected in that within this idiomatic schema, whereas the semantic meaning of N_1 is relevant, the semantic meaning of N_2 is implicitly more informative containing important message with regard to the speaker's judgement or attitude towards an individual or situation. The locative priority of N_1 over the informative proposition N_2 as $\{VN_1 + VN_2\}$ is supposed to introduce some background or contextual knowledge.

2. ABAC Structure

There are Chinese idiomatic constructions whose morpho-syntactic composition is unique in that in each of two subparts there is an identical morpheme, which can be illustrated as AB-AC; AB-CB or AB-CA. Idioms falling under this category are 见仁见智 (jiàn-rén-jiàn-zhì); 出尔反尔 (chū-ěr-fǎn-ěr); 相辅相成 (xiāng-fǔ-xiāng-chéng); 自卖自夸 (zì-mài-zì-kuā), etc. The importance of studying this specific morpho-syntactic pattern of Chinese idioms is reflected in the fact that among the twenty-nine types of partially-filled idiomatic structures which have been highlighted in the *Dictionary of Modern Chinese* as remarkably productive, the proportion of semi-specified idiomatic patterns that can be generalized as AB-AC or AB-CB represents nearly one fourth of the totality (Jian, 2001: 550-4). A very representative CIC of this ABAC structure type is {不 (no; not)…不…}, which has proved to be highly productive in modern Chinese with some twenty instantiated idiomatic expressions identified from the CMC, topping list of ABAC type CICs highlighted in the present study:

{半 (half)…半…} (12)6; {大 (greatly; big; of important proportion) …大…} (14); {没 (without)…没…} (6); {无 (without)…无…} (15); {有 (with)…有…} (13); {一 (one; single)… 一 …} (11); {多 (many)…多…} (7); {自 (by oneself)…自…} (13); {可 (worth)…可…} (7); and {尽 (exhaustively)…尽…} (4).

- {不 (no; not)…不 (no; not)…}

2.1 不知不觉 (bù-zhī-bù-jué) (829); 3.2 不屈不挠 (bù-qū-bù-náo) (260); 3.3 不折不扣 (bú-zhé-bú-kòu) (221); 3.4 不慌不忙 (bù-huāng-bù-máng) (225); 3.5 不声不响 (bù-shēng-bù-xiǎng) (193); 3.6 不伦不类 (bù-lún-bú-lèi) (120); 3.7 不明不白 (bù-míng-bù-bái) (109); 3.8 不三不四 (bù-sān-bù-sì) (86); 3.9 不闻不问 (bù-wén-bù-wèn) (85); 3.10 不偏不倚 (bù-piān-bù-yǐ) (68); 3.11不干不净 (bù-gān-bù-jìng) (18); 3.12 不上不 下 (bù-shàng-bù-xià) (9); 3.13 不蔓不枝 (bù-màn-bù-zhī) (1); 3.14 不破不立 (bù-pò-bù-lì) (10); 3.15不痛不痒 (bù-tòng-bù-yǎng) (37); 3.16不即不离 (bù-jì-bù-lí) (29); 3.17 不卑不亢 (bù-kēng-bù-kāng) (87); 3.18 不骄不躁 (bù-jiāo-bù-zào) (23); 3.19不理不睬 (bù-lǐ-bù-cǎi) (30); 3.20不吐不快 (bù-tǔ-bù-kuài) (15).

The statistical significance of the construction {不 (no; not)…不…} indicates that this morpho-syntactically symmetrical and phonetically balanced CIC has a wide application domain in modern Chinese. However, the fact that whereas Chinese native speakers may well explore this CIC in agreement with the contextual situations, foreign language-learners have found this idiomatic pattern rather elusive to imitate seems to suggest that its semantic

[6] Here, the number in brackets refers to the totality of idiomatic expressions detected from the CMC that share the same CIC.

and pragmatic properties which restrict its flexible use in actual contexts may be far more complex than those of a redundant negative structure that can be readily approached as [not A + not B]. A closer look at the twenty idiomatic expressions constructed within this pattern as shown above leads to the conclusion that the semantic relationship between the bipartite constituent of each item as [不A] + [不B] can be subsumed under three major categories whereby (1) A and B are semantically relevant or complementary as in 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.13, 3.15, 3.18, 3.19; (2) A and B are semantically contrastive or opposite like in 3.12, 3.16, 3.17; and (3) A and B are semantically resultant like in 3.14 and 3.20. In screening {不A不B} instances characterized by the semantic complementarity or relevance between A and B, an interesting linguistic phenomenon emerges in terms of the collocational feature that many A/B morpheme pair bears. That is, when the semantic meaning of the monosyllabic A and B is corresponding in the CIC of {不A 不 B}, it seems that they tend to be combined together to form a disyllabic character in modern Chinese.

   Within the three major categories established on account of the semantic relation between the two underspecified morphemes A and B within this construction, the first category of semantic complementarity has proved to be more complex than the other two, thus worth more detailed analysis. In an instantiated expression built on the pattern of {不A不B} where A and B are semantically complementary, there are at least three kinds of possibilities for this partially complementary nature to be extended at a higher level as that of the idiomatic expression. For example, in不知不觉 (3.1), 不折不扣 (3.2), 不慌不忙 (3.4), 不声不响 (3.5), 不偏不倚 (3.10), 不骄不躁 (3.18) and不痛不痒 (3.15), the semantic interference between {不A} and {不B} remains at a minimum given that the meaning of A and B are relatively independent and may be taken separately, thus the idiom may be safely approached as a simple reinforcement of its compositional parts: {不A 不B} = {不A} +{不B}. Meanwhile, in idioms or collocations like 不明不白 (3.7) and 不干不净 (3.11), the semantic complementarity between A and B begins to show a higher degree of dependency, where an isolating treatment of either part, as has proved to be feasible with the former type of {不A不B }, would necessarily undermine the semantic integrity of the idiom as a whole. In fact, here a proper interpretation of idioms of this sort can be illustrated as {不A 不B} = {half A} + {half B}.

   Lastly, there are {不A不B} type idioms like不伦不类 (3.6), 不三不四 (3.8) and 不蔓不枝 (3.13) where the semantic meaning of A and B exhibits such level of interactivity that when taken together within this particular morpho-syntactic pattern they border on a culturally conventionalized pragmatic meaning that can be least predictable when relying solely on the knowledge of their literal sense in a non-idiomatic textual environment. For example, in explaining不伦不类 (3.6), it should be noticed that the meaning of the idiom is not equal to a simple combination of {不伦 (without - match; peer)} and {不类 (without - type; class)}; in other words, the exploration of the CIC of {不A不B} by the noun morpheme pair 伦-类 is not meant to be emphatic or comprising as in the previous two cases, but largely pragmatic – when allocated within this pattern, 不伦不类 refers to one's disgraceful or scandalous behavior rather than a classifying practice as the literal meaning of this idiom would suggest. In this respect, the collocation of 三 (three) and 四 (four) within the construction of {不A不B} is even more illustrative than the former for the necessity of treating certain CICs as pre-constructed lexical patterns: without being well aware of the conventionality implied in this idiomatic pattern when it is used with this particular morpheme pair, one would hardly dare hazard a guess at the meaning of this idiom as referring to one's dubious or frivolous way of behave rather than the action of counting numbers. Thus, in the last case, the semantic relationship between the idiom as a whole and its integral parts may be expressed as {不A 不B} ≠ {不A} + {不B}.

IV Conclusion

The contribution made by the current study on an important subject of Chinese phraseology, i.e. the structural versatility of Chinese idioms, is two-fold: firstly, it is believed to be the first study of its kind in deploying corpus linguistic tools to retrieve quantitative relevant data from large-scale corpora; and the corpus data thus gathered has played a vital role in the identification and description of important linguistic phenomena with regard to the behaviour of Chinese idioms in naturally-occurring contexts. It seems that previous studies pursued within the field of Chinese phraseology or lexicology have been largely prescriptive and static, which have thus proved to be of very little help in bringing us any closer to the nature and particularity of Chinese idioms. The strictly-defined corpus-driven approach to the structural variability of Chinese idioms in the real use of the language hopes to bring the study of Chinese idioms in line with modern phraseology and lexicology as informed by the substantial use of corpus material in describing and analysing the linguistic properties of idiomatic expressions (Sinclair 1984; 1987 & 1991; Fernando, 1996; Moon, 1998; Hanks, 2003).

  Secondly, under the hypothesis that idioms and idiomatic expressions are subject to a special set of rules governing their linguistic behaviour, and thus can be reasonably systematized or further integrated into a model of the general patterning of each language (Moon, 1998: 150), more efforts have been made in the present study to situate the description and analysis of Chinese idiomatic structures within an entrenched framework that takes into account important linguistic factors serving as constraints on the morpho-syntactic configuration of Chinese idiomatic variants. Such a frame-theory-based research schema, as diametrically opposed to traditional single dimensional grammars, has proved to be prolific in that it has yielded us valuable insights into the rationale behind the seemingly idiosyncratic behaviour of Chinese idioms in naturally-occurring contexts, e.g. the projection of the semantic relation holding between the underlined morpheme pair onto the underspecified morpheme pair of a CIC, as shown in the idiomatic mould of {改 (change)…换 (change)…}; the pragmatic imbalance implicit in the paired focus construction of {有 (with)…无 (without)…}; and the semantic compositionality of double negative structure {不 (no; without)…不 (no; without)…}. It seems that the complexity of the phenomena of Chinese idioms is fully explainable only in terms of the correlated linguistic factors operating at different linguistic levels, such as syntactic, semantic and pragmatic. In this sense, the preliminary results of the present study, bearing on the structural variability of certain types of Chinese idioms from a largely descriptive perspective, would help point in the right direction for future research on the subject.

Reference

Atkins, B. T. S (1995) "Word meaning as a gestalt: a frame-semantic approach to lexical description", in Ravin, Yael (ed.) *Polysemy,* Oxford: OUP

Biber, Douglas (1993) "Co-occurrence patterns among collocations: a tool for corpus-based lexical knowledge acquisition", in *Computational Linguistics*, 19(3): 531-528

Biber, Douglas; Conrad, Susan & Reppen, Randi (1994) "Corpus-based approaches to issues in applied linguistics", in *Applied Linguistics* 15 (2): 169-189

Biber, Douglas; Conrad, Susan & Reppen, Randi (1998) *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: CUP

Bolinger, D (1977) "Idioms have relations", in *Forum Linguisticum,* 2/2, pp. 157-69

Brown, G. and Yule, G (1983) *Discourse Analysis*, Cambridge: CUP

Burkhanov, Igor (2003) "Pragmatic Specifications: Usage indications, labels, examples; dictionary of style; dictionary of collocations", in van Sterkenburg, Piet (ed.) *A Practical Guide to Lexicography*, London: John Benjamins Publishing

Fillmore, C. J. & Atkins, B. T. S. (1994) "Starting where the dictionaries stop: the challenge for computational lexicography", in *Computational Approaches to Lexicon*, Oxford: OUP, pp. 349-393

Fillmore, C. J. (1976) "The need for a frame semantics within linguistics", in *Statistical Methods in Linguistics*, pp. 5-29

Fillmore, C. J. (1985) "Frames and the semantics of understanding", in *Quaderni di Semantica,* 12, pp. 222-54

Fillmore, C. J. and O' Connor, M.C. (1988) "Regularity and idiomaticity in grammatical construction: the case of *let alone*", in *Language,* 64/3, pp. 501-38

Hanks, Patrick (2003) "Lexicography", in Mitkov, Ruslan (ed.) *The Oxford Handbook of Computational Linguistics*, Oxford: OUP, pp. 48-69

Ji, M (2007a) "Enhanced Phraseological Idiomaticity in Chinese Translational & Non-Translational Texts", paper presented at *Corpus Linguistics 2007*, Birmingham, UK

--- (2007b) "The use of language corpora in the study of Chinese idioms", paper presented at the *16th Postgraduate Conference in Linguistics*, Manchester

Kilgarriff, Adam & Rundell, Michael (2002) "Lexical profiling system and its lexicographic applications: a case study", in *proceedings of Euralex 2002*

Kilgarriff, Adam & Tugwell, David (2002) "Sketching words". In Correard, M. H. (ed.) *Lexicography and Natural Language Processing: a festschrift in hour of B. T. S. Atkins,* Euralex, pp. 125-137

Li, Ru Long (2002) 汉语的衍生方式及其演变, in *Journal of Normal University of Shi Jia Zhuang,* 5, pp. 68-76

Milner, G. M. (1969) "Quadripartite structures", in *Proverbium*, 14

Minsky, M. (1980) "A framework for representing knowledge", in D. Metzing (ed.), *Frame Conceptions and Text Understanding*, Berlin and New York: Walter de Gruyter, 1-25

Moon, Rosamund (1996) "Data, description and idioms in corpus lexicography", in Martin Gellerstam, et al (eds.) *Proceedings of Euralex 1996*, Goteborg: Goteborg University, pp. 245-256

Moon, Rosamund (1998) *Fixed Expressions and Idioms in English*, Oxford: OUP

Moon, Rosamund (1998) *Fixed Expressions and Idioms in English: a Corpus-based Approach,* (Oxford Studies in Lexicography and Lexicology), Oxford: OUP

Moon, Rosamund (1999) "Needles and haystacks, idioms and corpora: gaining insights into idioms, using corpus analysis", in Herbst, T. & Popp, K (eds.) *The Perfect Learners' Dictionary* (Lexicographica Series Maior 95) Tubingen: Max Niemeyer, pp. 265-281

Moon, Rosamund (2000) "Phraseology and early English dictionaries: the growth of tradition", in Heid, U. *et al* (eds.) *Proceedings of the Ninth Euralex International Congress, Euralex 2000*, Stuttgart: University of Stuttgart, pp. 507-516

Nagy, W. (1978) "Some non-idiom larger-than-word units in the lexicon", in D. Farkas, W. M. Jacobsen, and K. W. Todrys (eds.), *Papers from the Parasession on the Lexicon, Fourteenth Regional Meeting of the Chicago Linguistic Society*, 289-300

Ooi, Vincent B. Y. (2000) *Computer Corpus Lexicography*, Edinburgh: Edinburgh University Press

Pearson, Jennifer (1998) *Terms in Context,* New York: John Benjamin Publishing

Robert W. P. Luk, A. B. Y. Ng (1998) "Computer-assisted learning of Chinese idioms", in *Journal of Computer Assisted Learning*, 14(1), pp. 2-18

Ruhl, C.E. (1975) "Primary verbs", in A. Makkai and V. B.Makkai (eds.) *The First LACUS Forum,* Columbia, SC: Hornbeam Press, pp. 456-66

Ruhl, C.E. (1979) "Alleged idioms with *hit*", in W. Wolck and P.L. Garvin (eds.) *The Fifth LACUS Forum*, Columbia, SC: Hornbeam Press, pp. 93-107

Shei, Chris (2005) "Phrasal Units in Chinese Collocation, Idioms, Extended Lexical Units, Lexical Chains", University of Swansea URL: http://www.swan.ac.uk/cals/staff/shei/publication/PhrasalChinese.htm

Sinclair, John (2003) "Corpora for lexicography" and "Corpus Processing", in van Sterkenburg, Piet (ed.) *A Practical Guide to Lexicography*, London: John Benjamins Publishing

Teubert, Wolfgang (2001) "Corpus linguistics and lexicography", in I*nternational Journal of Corpus Linguistics*, vol.6

Wang, J (2001) "Recent Progress in Corpus Linguistics in China", in *International Journal of Corpus Linguistics*

Xue, Shiqi (2003) "Chinese Lexicography: past and present", in R. R. K. Hartmann (ed.) *Lexicography: Critical Concepts*, New York: Routledge