A Corpus-based Study of the Colligational Features of China English

LIANG Jianli

School of Foreign Languages, Huizhou University, China

Received: June 23, 2021 Accepted: July 21, 2021 Published: November 30, 2021

To cite this article: LIANG Jianli. (2021). A Corpus-Based Study of the Colligational Features of China English. *Asia-Pacific Journal of Humanities and Social Sciences*, 01:3, 094-113, DOI: 10.53789/j.1653-0465.2021.0103.011.p

To link to this article: https://doi.org/10.53789/ j.1653-0465.2021.0103.011.p

The research is supported by Education Sciences Planning Foundation of Department of Education of Guangdong Province (No. 2018GXJK185), and the Educational Research Fund of Huizhou University (No. JG2019015).

Abstract: The paper shows the findings of a study of China English (alternatively called Chinese English, both referred to as CE in this paper) based on the China English Corpus. The paper firstly introduces the research background of CE collocations. Three distinct features of collocations are identified and interpreted in the collocational theory framework. It reveals extensive collocational variations across written CE, thus complementing previous observations about deviations in English used by CE users and other varieties of English. Some issues concerning implications and future research scales are also discussed. The study shows how features of CE can be examined from the perspective of reoccurring patterns using the corpus tool, which enables us to deal with overt and covert patterns in a large amount in a repeatable way.

Keywords: China English Corpus; China English; colligation

Notes on the contributor: LIANG Jianli is an associate professor of the School of Foreign Languages, Huizhou University. Her research interest lies in corpus linguistics and language teacher education.

1. Introduction

As Bolton & Botha (2015: 169) state: "The role of English in Chinese society today cannot be considered in isolation from the sociolinguistic background, as well as the social and political context of contemporary Chinese society".

The past decades have seen unprecedented growth in the modes and diversity of language contact, partly due to forces of globalization and partly due to increasingly convenient access to the internet via all kinds of user-friendly smartphones and other mobile electronic communication devices (e.g., Blommaert 2010; Crystal 2001, among others). Mobile phones connected to the internet have largely redefined patterns of human interaction as well as modes of communication, which in turn help accelerate the processes of globalization. Under these cir-

cumstances, diversity and change in English language usages should also be recognized because "it is just as likely that the course of the English language is going to be influenced by those who speak it as a non-native language as by those who speak it as a mother tongue" (Crystal 2012: 167).

According to statistics published by International Telecommunication Union (ITU), a United Nations specialized agency for information and communication technologies, in 2013, over 2.7 billion people were using the internet, which corresponded with approximately 39% of the world's population (The World in 2013: ICT Facts and Figures 2014: 2). Advancement in ICT mediated by computers and smartphones and the increasing popularity of the internet has dramatically facilitated communication over a distance connecting people from different parts of the world.

Such forces of globalization and the convenient access to a wide range of information via the internet have greatly favored the development of certain languages as *lingua franca*, especially English (Haberland 2013). In an age of digital information, English knowledge is required to gain access to 'breaking news' broadcast in English. For instance, in the wake of the Edward Snowden news story from June 2013, or the two doomed Malaysian Airlines aircraft MH370 (March 2014) and MH17 (July 2014), first-hand information tended to be made available in English. Internet users who wish to quickly find out and are keen to keep up with such news stories would have no access to such information if they did not know English. In addition, knowledge of English in following international news provides bilingual users of English with multiple perspectives as seen through different TV news agencies across different vantage points and contexts. The will to participate in events that attract individuals' attention worldwide has reached a record-high level. For example, after the 1-7 Brazil vs. Germany match on July 9th, 2014, some 66 million people around the world created (i.e. posted, discussed, or gave comments on) a total of over 0.2 billion postings on the Facebook pages 4 hours after the match (Retrieved from www.cnet.com). The active and instant mode of communication online has provided a strong incentive for bilingual users of English to get involved in meaning-making activities using new technologies.

This helps explain why in many parts of the world, for local bilingual users of English, English enjoys an important status of a lingua franca, if not the lingua franca of choice, and that the numbers of bilingual users of English regardless of their first language(s) keep expanding. As Haberland and Mortensen (2012: 1) put it, "English is *not* spoken in every corner of the world, just in more places than any other language ever before" (emphasis in original).

Over half a century ago, in his book *Bilingualism as a World Problem*, Mackey (1967: 18) proposed that the key reason for educated people to learn English is "to be able to seek the knowledge they need in one or more of the languages in which most of the world's knowledge is available". However, with the rapid development of information and communication technology in the past few decades, the educated elite are no longer affected; rather, blue-collar workers in non-English-speaking countries who wish to be informed of and connected to other people may perceive a strong need to use English to understand what is going on elsewhere in the world and communicate with people who do not share their first or usual language. Whatever their first language, educated or otherwise, people all over the world are connected; they are willing and readily facilitated to surf the internet to consume a wide range of information, from browsing breaking news and funny video clips to searching for options of entertainment and leisure activities such as popular restaurants and internet games with the most 'Likes'. The internet also provides a platform for social interaction, such as chatting with friends at e-forums, sharing opinions in blogs, and giving 'Likes' on Facebook pages. Such out-of-class, internet-mediated activities have become an integral part of our everyday life; the wide-ranging need for information in English plus a strong mo-

tivation to be connected with people from different language backgrounds has given English a special status, a lingua franca for international communication. This need for communication continues to maintain English as the language that people cannot afford not to learn and use nowadays.

Commonly referred to as 'the world's lingua Franca, English plays a vital role in the globalization-in-a-multilingual-world movement, which results in an "English language complex" (Mesthrie & Bhatt 2008: 1-3). "We're all 'global' now, and need to use the first truly universal language, whether we are business people, politicians, teachers, tourists, or terrorists" (McArthur 2003: 54). "What happens, linguistically, when the numbers of the human race use a technology enabling any of them to be in routine contact with anyone else?" (Crystal 2001: 5).

It is already a fact that the spread of English around the world has been, and continues to be, both rapid and unpredictable. One consequence of colonization (in the past) and globalization (at present) is the general spread of English on the one hand and the inclusion of English language teaching in the national or local education curriculum on the other (Jenkins 2007). Being an increasingly important and active player in global political and economic affairs, China, with the largest numbers of users of lingua franca English in the world, is one such nation where the above-mentioned forces of globalization and the growing popularity of the internet are played out. Users of English in China face many challenges. These include a) English has become an important part of compulsory education, lasting for 13 to 15 years from primary to tertiary education depending on the onset year (Grade 3 or Grade 1) (Adamson 2004, 2014); and b) the objective need as well as subjective wish to be increasingly engaged in the English-dominant internet makes it necessary for millions of Chinese nationals to use English more or less regularly, both as consumers of information and as agents interacting with people from different language backgrounds (Bolton & Botha 2015). Such conditions have already greatly influenced the language ecologies in China today, especially the role it plays for the nation as well as for individual citizens. Politically, English has become one very important tool for the nation to build up her image and articulate her diplomatic stance in international, especially bilateral issues. Domestically, multilingual users of English now enjoy more and more global horizons via English. In short, the sociopolitical role of China in the world, together with the rapidly increasing use of English-oriented information and communications technologies, constitutes the background and setting for Chinese people to use English as part of their daily life activities.

The process and place of the learning of English in the Mainland Chinese education system have not changed much, given that classroom-based teaching and learning has always been the key setting where English is acquired. On the other hand, the range of out-of-class social contexts where English is naturally used is fast expanding, even though the language learning context remains more or less the same in the last 30 years: children are expected to learn Putonghua, the national language, from kindergarten, while the onset age for learning English begins at Grade 3 (age 9) or Grade 1 (age 6) depending on the region.

English, as used in Mainland China, is the variety we concentrate on in this paper. More specifically, our main focus is on the features of English as found in the written outputs gathered from Han Chinese authors in Mainland China. We will briefly review debates about CE, the possibilities of codifying CE patterns, and other recent developments from a World Englishes perspective. Then we will present and describe a detailed account of several salient structures and collocation patterns which reveals the overt and covert collocational patterns.

2. Literature Review

Over the last three decades, scholars promoting the research and codification of CE have worked towards establishing a better understanding of the features of CE. In a rough-and-ready way, the relatively brief progress of CE research can be seen as falling into the following three categories:

a. the existence of CE and the issue of terminology (e.g.: Cheng 1992; Evans 2011; Ge 1980; Gui 1988; He & Li 2009; Huang 1988; Jiang & Du 2003; Kirkpatrick & Xu 2002; Niu & Wolff 2003; Wang & Ma 2002; Xu 2010; Zhuang 2000);
b. attitudes towards CE (e.g.: He & Li 2009; Hu 2004, 2005; Pan & Seargeant 2012; Xu 2010);
c. linguistic features of CE (Bolton 2003; Chen 2010; Gao 2008; Yu 2009; Kirkpatrick 2007; Li D. C. S. 2002; Sun 2011; Xu 2010).

Some comprehensive bibliographies (Adamson, Bolton, Lam & Tong2002; Bolton, Botha & Zhang 2015) provide details of literature on CE studies.

What is clear from current research could be seen from the following two quotes:

a. …there is some evidence that 'China English' is gradually emerging, following its natural path of development, although it is quite impossible to list all the linguistic features of 'China English' exhaustively at the moment for several reasons, such as insufficient research. Therefore, more research is needed to identify salient linguistic features of 'China English' as found in the popular usage patterns of the majority of speakers and writers of 'China English', in both formal and informal contexts of social interaction. (He & Li 2009: 74);
b. The current popularity of English in China is unprecedented and has been fuelled by the recent political and social development of Chinese society (Bolton & Graddol 2012: 3)…there is an evident need to carry out more field-based sociolinguistic research. (ibid: 7)

3. The Study

3.1 Data and tool

The target corpus data consists of 37 million words of English collected from Mainland Chinese users and is referred to as CEC. British National Corpus BNCweb (CQP-edition) (referred to as BNC) is used as the reference corpus (see Table 1):

Target corpus	Reference corpus
China English Corpus (CEC)	British National Corpus (BNC)

Table 1: Target corpus and reference corpus

CEC

CEC (Li, W. Z. 2010) is the largest corpus currently available that represents the English used by educated Chinese people (main people receiving the undergraduate level of education or above). It is large enough (37,

470, 040 word tokens) to generate representative results. This corpus was constructed explicitly to investigate CE linguistic features. To this end, the data collected seek to meet the criteria of genre balance (e.g., written events include educational, leisure, natural and social sciences, and business studies and communication, etc.) and a range of domains (see Table 2 below). The writers of texts in CEC share similar linguistic backgrounds in Mainland China.

As a corpus, CEC is a collection of samples of written English from a wide range of sources in China which, by design, represent a broad cross-section of China English. It is an output of an academic research project funded by the government of China and carried out by a team of professional researchers (university faculties) led by Professor Li Wenzhong in Henan Normal University in Mainland China since 2001. It consists of a total of 17, 534 texts collected according to the genres of BNC (written part).

For example, CEC includes extracts from regional and national newspapers published in Mainland China, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published information bulletins, published theses and papers, among many other kinds of written text. The comparable nature of these two corpora in terms of domain classification is listed in Table 2 below:

	BNC domains	**CEC domains**
1	natural & pure science	natural science
2	applied science	applied science
3	social science	social science
4	world affairs	world affairs
5	commerce & finance	economics
6	arts	arts
7	belief & thoughts	beliefs & thoughts
8	leisure	leisure
9	literature	literature

Table 2: Domains of CEC and BNC

(Source: Aston & Burnard, 1998: 29; W. Z. Li 2005)

Written English represents one consistent type of English output of educated CE users; for obvious reasons, written data, which is easier to obtain in useful quantities than spoken data, serves as a convenient, ready-to-use database for corpus-based linguistic investigation. The readership of these written English data is two-fold: first, those Mainland readers who have the ability and will to access information disseminated in multiple channels and languages, including English; second, readers who are driven by a desire to improve their English by giving themselves additional opportunities and exposure to that target language more or less regularly. Foreigners with no Chinese language background could only access information (e.g., through English newspapers in China) or through web pages whose contents, in the Chinese context, are produced by CE users. Today, there is an increasing number of educated people in China who would like to read a variety of texts written in English. Therefore, there is a need for the written sources to be comprehensible and to convey practical information with the two kinds of readership — Chinese and non-Chinese — as target consumers, which is the key function that these home-grown written English outputs are serving. Produced with that purposeful goal and target readers in

mind, the components of CEC are thus ideal sources for the current investigation.

As for the representativeness of CEC data, this study follows the same premises that previous pioneering studies adhered to regarding the representativeness of corpus data. Leech, for example, argues that "there is a scale of representativity" (Leech 2007: 144) in data. Similarly, McEnery and Hardie state that "the measures of balance and representativeness are matters of degree." (McEnery & Hardie 2012: 10). Thus the representativeness of CEC should be calibrated by the extent to which it truly reflects the preferred linguistic features of Mainland Chinese users of English collectively as a whole. Texts in CEC represent CE in two ways: 'mainstream' and 'general'. CEC represents mainstream CE because it includes articles from mainstream newspapers which are categorized into ten components, whose sizes and sub-categories are listed in Table3.The texts were sampled from different genres in more than 30 domains, including journal articles, academic essays (wide coverage of topics/disciplines), news reports (reportages and reviews), editorials, public relations documents from public and private organizations, etc. These genres are all comparable to those in BNC.

Topics	No. of files	Tokens	Types	Topics
Natural science	1, 324	3, 174, 622	110, 980	mathematics, physics, chemistry, biology, astronomy
Applied science	546	953, 951	45, 425	engineering, communications, technology, computing, energy, transport, aviation
Social science	4, 494	1, 6079, 207	248, 045	sociology, geography, anthropology, medicine, psychology, law, education, linguistics
World affairs	2, 358	2, 535, 967	57, 448	history, government, politics, military, archaeology, current events
Economics	1, 553	2, 003, 378	61, 934	business, finance, agriculture, industry, third industry, employment
Arts	1, 944	1, 346, 122	55, 524	visual arts, calligraphy, brushwork, Chinese Wushu (martial arts), architecture, performing arts, media studies, carvings
Beliefs & thoughts	998	562, 475	27, 772	religion, philosophy, folklore
Leisure	2, 599	1, 444, 268	62, 133	food, travel, fashion, sport, household antiques, hobbies, gardening
Literature	283	8, 653, 464	117, 561	fiction, prose, drama scripts, classics

Table 3: Domains and description of the CEC data

* **Note:**

- Tokens and types are calculated with the help of the corpus tool AntConc (Version 3.5.8) (Anthony 2019).
- Firstly, using AntCoc, the names of authors of articles included were checked to establish whether they are Chinese. Chinese names follow a format (e.g., ZHANG Kexin), while names of people not from China would look different (e.g., Matsuda Aya, which is not recognizable by the pinyin system). This precautionary check is to ensure that all data represent CE.

Despite some possible shortcomings (e.g., lack of spoken data), the readily prepared CE data still serve as one of the best sources for researching CE features for two main reasons: representativeness in terms of size and authenticity in terms of function.

a. Representativeness in terms of size

There are 37, 470, 040 word tokens in the CEC. Information about frequency is one of the most obvious benefits that a corpus can provide, which cannot be provided by any other mode of linguistic analysis. Then, in a corpus of the size of the CEC, the information about frequency is more convincing than intuition-based theorizing. The importance of frequency information is also endorsed by the mathematical theory, which calculates

words (or lexemes) according to the patterns in which they appear and generates hypotheses from it. The mathematical theory in support of this method is Zipf's Law of word distributions. Zipf (1935, 1949) held that in any language corpus, the frequency of any word is inversely proportional to its rank in the ranking table of frequency, and the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. Zipf's Law gives more heavily weighted importance to the most frequent words than would be expected according to the normal distribution in language. In other words, the larger the data, the more suitable it is in terms of obtaining reliable and representative frequency information.

b. Authenticity in terms of function

As discussed earlier, corpus linguistics focuses on authenticity. Some may argue that written articles are not as authentic as daily face-to-face conversation. However, as indicated earlier, there are millions of competent users of English in China (as a result of the education they received and the computer networks that have evolved as platforms with which CE users communicate with the rest of the world). However, English is not commonly used between Mainland Chinese when there are no English-speaking people around, so the collection of written and spoken data from spontaneous situations is difficult and not quite feasible, at least not at present. In this light, CEC could be seen as representing how Chinese users of English write when conveying information to target readers they hope to reach at home or abroad. The authors who wrote the articles in question may have overseas experience and speak 'native-like English; however, even here, CE features will be found if they are indeed CE features. The ultimate purpose that each article fulfills is to accomplish the communicative function in written form, which is authentic too in every sense of the term.

A contrastive comparison was made between the target corpus (CEC) and a reference corpus (BNC). As Leech (2002) argues, a reference corpus is important in any empirical investigation because it serves as a benchmark and yardstick and provides more comprehensive information about the linguistic features of the language under investigation.

BNC

The British National Corpus (BNC) (Leech, Rayson, & Wilson 2001) is a 100-million-word structured collection of spoken and written texts. The corpus was compiled by a consortium of universities, publishers, and the British government in the 1990s, to be representative of the spoken and written English used by British people toward the end of the 20th century, with written data amounting to around 90, 000, 000 tokens. The front-end interface of the BNC is available on the BNCweb (http://bncweb.lancs.ac.uk/). After registration, users can access the BNCweb according to their search targets. The search in the case of this study was limited to written form in the BNCweb, to ensure that it is comparable in size and genre to the CEC.

Tool

Whilst there are lots of computer-based tools and web-based tools available in corpus studies, we choose the commonly-used free corpus toolkit AntConc (Anthony 2019) in our research.

3.2 Research framework

To identify possible patterns that involve words occurring next to each other we were unaware of before, we need to learn relatively fixed, largely pre-defined sequences of words in more detail. Colligational patterns are thus chosen to be the study focus. Colligation is the co-occurrence of words with grammatical choices (e.g., adjective + noun) (Sinclair 2004: 174). First, we tagged CEC using free software called *Lemmatizer* (Li W. Z. &Liang M.C. 2010, available online). To conduct our search within workable limits, the modifying patterns of nouns are chosen as the target.

Cross-linguistically, the modifying structures of nouns are very diverse, but the major constituents are nonetheless discernible. According to Leech, Deuchar & Hoogenraad (2006: 71), there are two kinds of modifiers for nouns, namely, premodifiers and postmodifiers. The possible premodifiers and postmodifiers of noun phrases are shown in Tables 4 and 5:

Type of premodifier	Example
determiners	the/an apple

continue

Type of premodifier	Example
enumerators	three apples
adjectives	big apples
nouns	gold ring
genitive phrases	Tom's problem
adverbs	quite a noise
other categories	awfully bad weather (adjective phrase) kind-hearted man (compound words) grated cheese (past-participle of verbs) a working mother (present participle form of verbs)

Table 4: Possible premodifiers of noun phrases (Leech et al. 2006: 71-72)

Type of postmodifier	Example
prepositional phrases	the best day of my life
relative clauses	a man that I admire
Other categories	the room upstairs (adverb) something nasty (adjective)

Table 5: Possible postmodifiers of noun phrases (ibid.)

At this stage, words are being analyzed in word classes, so their analysis is colligational by nature. The colligation determiners + nouns are not studied in this study because determiners are not comparable in lingua-cultural significance concerning Chinese. 13 high-frequency nouns are chosen to be the target node words for our study, whose information of occurrences in contrast to each other was summarized in Figure 1.

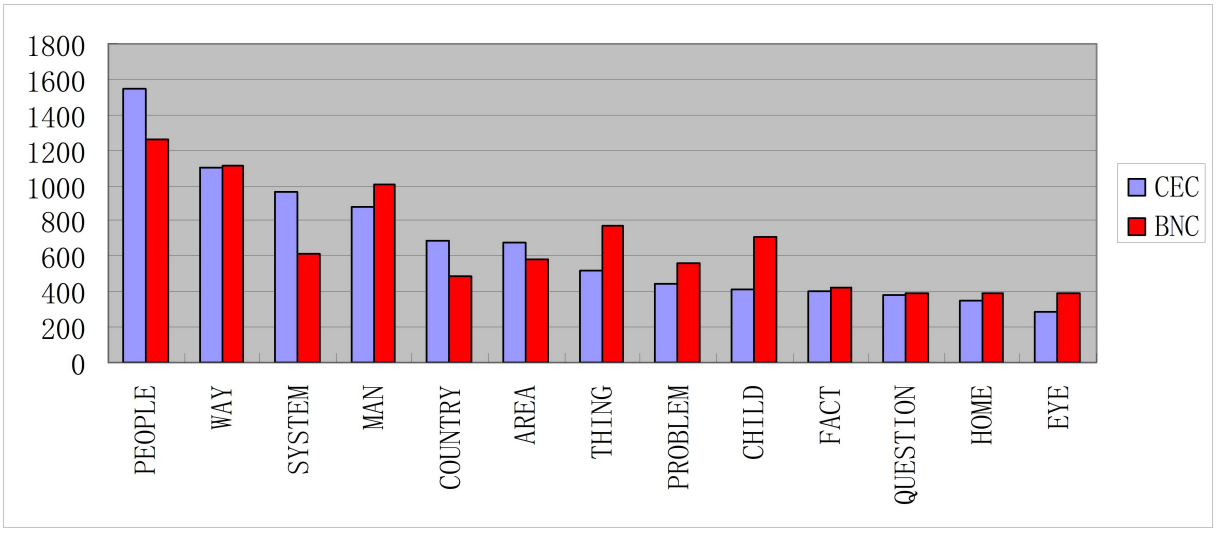

Figure 1: Comparative frequency data for target node words (per million words)

* Lemmatized forms are sometimes written in upper case. For example, the verb lemma WALK consists of the words walk, walked, walking, and walks. In this paper, words in an upper case like this are lemmas; others in lower case are word types. In language that can be observed, words are in the form of types (e.g., *is, am, are, was, were, been*, and *being* are all word types of the same lemma BE).

4. Results and Discussions

4.1 Enumerators as premodifiers in noun phrases

Broadly speaking, an enumerator consists of one or more words denoting a cardinal or ordinal number (e.g., 'one', 'five', 'double', 'the twelfth'). There is extensive evidence in the CE corpus data showing a clear preference among CE users for using enumerators (Table 6).

No.	CE collocations	Chinese	Pinyin
1	One Child Policy	計劃生育政策	jìhuà shēngyù zhèngcè
2	One Country Two System	一國兩制	yīguó liǎngzhì
3	One World One Dream	同一個世界 同一個夢想	tongyīgè shìjiè tongyīgè mèngxiǎng
4	One-China Principle	一個中國原則	yīgè zhōngguó yuánzē
5	Two Guarantees	兩個確保	liǎnggè quèbǎo
6	Double mugging (i.e., snatch and rob, generally referred to in English newspapers in China as 'Two Robbery')	雙搶	shuāngqiǎng
7	Two-State Theory	兩個中國	liǎnggè zhōngguó
8	Two-Way Investment	雙向投資	shuāngxiàng tóuzī
9	Three Antis Campaign	三反活動	sānfǎn huódòng
10	Three Direct Links	三通	sāntōng
11	Three Gorges Dam/Hydropower Station	三峽工程/水壩	sānxiá gōngchéng
12	Three Gorges Water Conservation Project	三峽水利工程計劃	sānxiá shǔilì gōngchéng
13	Three Kingdoms/Period	三國	sānguó
14	Three Public Consumptions	三公	sāngōng
15	Three Represents	三個代表	sāngè dàibiǎo
16	Three Worlds	三界	sānjiè
17	Three-Tier Rural Health Care Service Network	三級保障農村醫療服務系統	sānjí bǎozhàng nóngcūn yīliáo fúwù xìtǒng
18	Four Books	四書	sìshū
19	Four Heavenly Kings	四大天王	sìdà tiānwáng
20	The Four Beauties	四大美人	sìdà měirén
21	The Four Duty Gods	四大金剛	sìdà jīngāng
22	Five Classics	五經	wǔjīng

continue

No.	CE collocations	Chinese	Pinyin
23	Five Friendlies	福娃	fúwá
24	Five Generations Under One Roof	五代同堂	wǔdài tóngtáng
25	Five Holy/Sacred Mountains	五嶽	wǔyùe
26	The Five Dynasties (Period)	五代	wǔdài
27	The Five Elements Mountain	五行山	wǔxíngshān
28	The Five Principles Of Peaceful Coexistence	和平共處五項原則	hépíng gòngchú wǔxiàng yuánzē
29	Six Harmonies	六和	liùhé
30	Eight Diagrams/Trigrams	八卦	bāguà
31	Eight Model Plays	八個樣板戲	bāgè yàngbǎnxì
32	Eight Vajrapanis	八金剛	bājīngāng
33	The Eight Allied Forces	八國聯軍	bāguó liánjūn
34	The Eight Arrays	八陣圖	bāzhèntú
35	The Eight Immortals (Crossing The Sea)	八仙(過海)	bāxiān
36	Ten Kingdoms	十國	shíguó
37	The Twelfth Five-Year Plan /Program	第十二個五年計劃	dìshí' èrgèwǔniánjìhuà
38	Eighteen Guardians/Defenders	十八羅漢	shíbā lúohàn

Table 6: Examples of enumerators as premodifiers in CE noun phrases

In terms of form, the fact that enumerators function as premodifiers of noun phrases is not surprising. However, their manifestation in CE seems distinct in its way. In Chinese expressions like those in Table 6, the use of numbers symbolizes seriousness and carries an unmistakable connotation of authority. Most of the usages are based on the Chinese counterpart and are thus translations in CE. However, although the Chinese for One Child Policy does not include any enumerator, CE speakers still prefer the current way of expressing this idea.

These examples show that, even when people have some other rhetorical choice at hand, they still favor using enumerators to modify the noun, probably out of a concern for preserving its formality. This leads us to predict that whenever some kind of political idea is to be expressed and implemented, one could reasonably expect the use of enumerator + noun as a favored or preferred rhetorical strategy.

4.2 A preference for noun pre-modifiers

As we have seen, nouns and adjectives are both common pre-modifiers of nouns, but their preferred use by CE speakers is another issue to consider. It is found that Chinese speakers tend to use more nouns than adjectives as pre-modifiers in constructing noun phrases in their social interaction with others in English, and this is supported by corpus investigation in terms of frequency. Collocation results from the KWIC search (span=L-1 left one from the node word) are stored and tagged using the same tagset system as the BNC, namely the CLAWS 5 Tag-

set. The tagged words were then sorted in an Excel spreadsheet and counted. After calculating the nouns and adjectives for the 13 target lemmas, frequency information per million words is compiled. This standardized frequency information is comparable because the tokens of each corpus do not influence them.

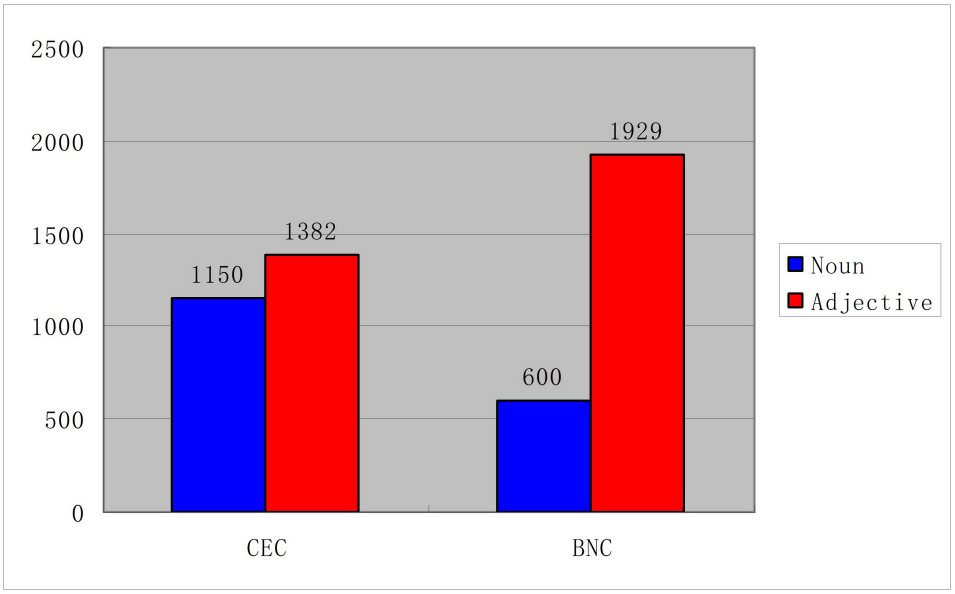

Figure 2: Contrasting the proportions of nouns and adjectives as premodifiers (per million words)

Overuse of the colligation *N* + *N*, the form usually referred to as a nominal compound, is an important and useful modification pattern in Mandarin Chinese (and most of the Chinese languages/dialects like Cantonese, Hokkien, and Hakka). The most influential description of this structure is provided in Li & Thompson (1981), which lists 21 types of colligation *N* + *N* in Chinese (Table 7).

Type	Example	CE *	BrE#
N1 denotes the place where N2 is located	臺燈	table lamp	table lamp
N1 denotes the place where N2 is applied	牙膏	toothpaste	toothpaste
N2 is used for N1	馬房	stable	stable
N2 denotes a protective device against N1	太陽眼鏡	sun glasses	sun glasses
N1 and N2 are parallel	國家	country	country
N2 denotes a product of N1	蠶絲	silk	silk
N2 denotes a place where N1 is sold	百貨公司	department store	department store
N2 denotes a disease of N1	心臟病	heart disease	heart disease
N1 denotes the time for N2	春天	spring	spring
N1 is the source of energy of N2	汽車	car	car
N2 is a component of N1	雞毛	feather	feather
N2 is a source of N1	煤礦	coal mine	coal mine

continue

Type	Example	CE *	BrE#
N1 denotes a proper name for N2, which may be a location, an organization, an institute, or a structure	北京大學	Peking university	Peking university
N2 denotes a unit of N1	政府機關	government organization	bureau
N2 denotes a piece of equipment used in a sport, N1	籃球框	basketball ring	basketball hoop
N2 is caused by N1	油漬	oil stain	stain
N2 denotes a container for N1	書包	schoolbag	schoolbag
N2 is made of N1	大理石地板	marbling floor	[not listed]
N1 is a metaphorical description of N2	龍船	dragon boat	dragon boat
N2 is an employee or an officer of N1	大學校長	university president	vice chancellor
N2 denotes a person who sells or delivers N1	鹽商	salt merchant	[not listed]

Table 7: Types of nominal compounds in Mandarin (Li & Thomson 1981: 49–53)

* Source: Chinese–English Dictionary (Wu 2010)

#Source: Cambridge English Dictionary Online(http://dictionary.cambridge.org/)

We can by no means assume that the 21 types of colligation *N* + *N* in Table 7 constitute an exhaustive categorization of the Chinese *N* + *N* pattern. The important thing to note here, however, is that the linguistic preference in Chinese may help us to better understand the more commonly used *N* + *N* patterns in CE. CE users link nouns and nouns together to form a nominal compound, with the effect of designating an object with a name, a productive and creative process which is rooted in their first language, and which by design expresses their indigenous worldview (e.g. the morpheme–for–morpheme translation of government organization is preferred to the more opaque bureau). And, the minor difference in some CE and BrE translations in Table 7 is further indicative of differences in CE as opposed to BrE norms (e.g. the high–frequency word *ring,* as in *basketball ring*, is preferred to the more obscure word *hoop*, as in *basketball hoop*).

4.3 Genitive phrases as pre–modifiers of nouns

Genitive phrases are acceptable and frequently used in both CE and BrE. The results of the colligation *N's* + *N* show that CE has the inner circle usage as its core, but is colored with characteristic features of the Chinese mindset.

In this section of the analysis, the first step involves a search for the structure: *N's* + *N* (of the 13 target lemmas), after which comparisons were made using two parameters: frequency and diversity. The frequency information of *N's* + *N* from the CEC and the BNC is similar, reflecting the fact that the colligation patterns involving nouns with a genitive case premodifier are largely shared by Chinese and British users of English —probably an extension of the meaning–making potential of a similar colligation or structure irrespective of their first languages.

Two features distinct to CE have emerged, the first of which is the variants and deviations in the use of people's as a premodifier, as shown in the following examples found in the earlier phase of the investigation (Table 8).

People's + N
people's area
people's thing
people's problem
people's way

Table 8: Examples of using people's as a pre-modifier in CEC

It was then decided to set *people's* as a target feature for closer scrutiny, partly because in my observation, the word "people" (in the Chinese language) has special importance to Chinese Mainlanders.

For example, the government (local or national) is termed 人民政府 (people's government); the federal bank is called 人民銀行(people's bank); one key newspaper is named 人民日報 (People Daily); all textbooks for primary to high school students are published by 人民教育出版社 (People's Education Press). Similar cases are not found in countries elsewhere. Further investigation on this structure has revealed more features.

One distinctive example is seen in the two-word cluster of *people's + ?*. The high-frequency occurrences of collocations are summarized below (Table 9).

Rank		Collocation	Tokens
1	people's	lives	174
2	people's	life	142
3	people's	mind	99
4	people's	attention	72
5	people's	health	64
6	people's	cognition	53
7	people's	understanding	49
8	people's	interest	43
9	people's	right	38
10	people's	eyes	35
11	people's	awareness	34
12	people's	attitude	33
13	people's	way	33
14	people's	experience	31
15	people's	perception	26
16	people's	desire	24
17	people's	thinking	21
18	people's	use	19
19	people's	consciousness	16
20	people's	cultural	11

Table 9: Top 20 two-word clusters modified by people's in CEC

Longer clusters such as *people's living standard*s, *people's daily lives*, and *people's conceptual systems* were also observed. These findings suggest that CE speakers are using the colligation N's + N in a pattern that differs from at least one inner circle variety—BrE. The use of people in the patterns of these sentences may also imply that CE conceives of human beings principally as part of a broader collective rather than an isolated individual. This is reflected in the use of the genitive case structure people's + ?in all appropriate cases, whereas British people use other words such as man, person, and the like to express essentially the same denotation. The difference is that in Chinese language, people (人民 totally different from 人們) carries a favorable connotation, while in BrE this word seems more neutral.

The second feature found in this part of the analysis is that CE speakers tend to underuse the colligation N's + N of. Examples of this pattern are shown in Figure 2 below. One kind of example is excluded here since they do not fit into the purpose of our investigation; they are institutional names such as *the People's Republic of China, People's Bank of China,* etc. It is common sense that thousands of these proper names would occur in CE data, yet they do not tell us much about what we hope to prove or disprove.

people's ability of recognizing the world. Image sche
people's acceptance of the letters is whether or not
people's acquirement of its merits. Besides, the inte
people's activities of cognition and mental experienc
people's anticipation of the event. Xi wished the U.S
people's assessment of the social significance of the
people's awareness of conservation of cultural and na
people's capability of bearing also varies from each
people's changes of taste and create new entertainmen
people's choice of health care services, he added. A
people's concept of fertility and in lowering the bir
people's conference of Beijing took the lead in prohi
people's confusion of identification in the white-dom
people's consciousness of the importance of marine pr
people's custom of dragon-boat racing during the Drag
people's definitions of the same term. However, there
people's demand of clearly keeping the eternal moment
people's descriptions of objects in the surrounding w
people's disobedience of the traffic rules, the unsci
people's dream of flying freely in the sky, has been
people's enjoyment of lanterns in the Song Dynasty (9
people's evaluation of other people's performance (Be
people's exercise of the right to be masters of the s
people's experience of the world and the way they per
people's feeling of devoid and absurdity are embodied

Figure 2: Examples of the pattern people's + N of in CEC

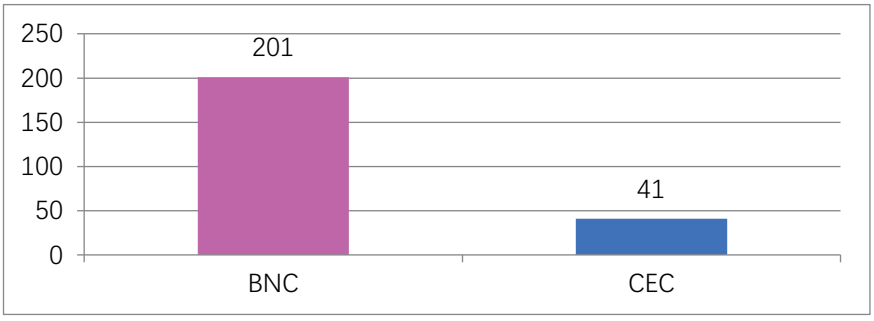

Figure 3: Contrasting the proportions of the colligation *N's + N of* in CEC and BNC(frequency per million words)

This colligational structure is constituted by both a premodifier and a postmodifier, and it does not, therefore, correspond with the Chinese way of expressing ideas. Han Chinese 'dialects' or languages are structured

according to fixed word order, namely the modifier-before-head sequence (Tai 1985; Ho 1993; Hu, W. 1995).

The genitive case in Tom's parallels that of 湯姆的 (tāngmǔ dē) in Chinese. Thus the -'s may be taken for granted (consciously or unconsciously) as a device for turning the modifier into an adjective, as is the case in Mandarin Chinese. However, the postmodifier does not have a functional counterpart in Chinese, so it is likely to be used less confidently by CE users in colligations such as N's + N of.

4.4 Post-modifying prepositional phrases

Prepositional phrases as post-modifiers of nouns occur with similar frequency in CEC and BNC. The most frequently used preposition in English is of. This study thus uses it as the target of the investigation.

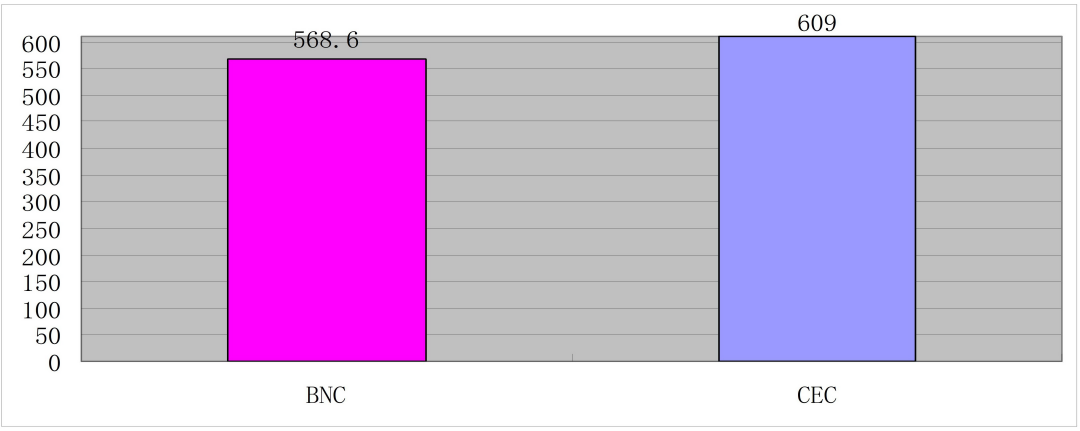

Figure 4: Contrasting the colligation *N* + *of* (frequency per million words)

4.5 Post-modifying relative clauses

Since it is complicated to obtain a finite description of a relative clause, this study chooses the most typical one, *that clause,* for sample, so the colligation under investigation is *N* + *that clause*.

Figure 5: Contrasting the colligation *N* + *that* (frequency per million words)

One possible problem was encountered when studying this pattern. Because in some sentences (e.g., He would like to give the girl that doll.), the word 'that' functions as a determiner rather than a relative clause

marker. Some extra work was done to ensure that Figure 5 was not skewed by these sentences. It was found that in the data observed, and this problem does not seem to affect the reliability of Figure 5. First, the search for this colligation generated a 222, 597−token result in BNC and a 60, 043−token result in CEC. Then, manual observations were taken to observe the first word to the right (R−1) of the pattern *N* + *that*. Not a single example, as within the limit of manual observation, was found in the five random pages of KWIC in both corpora.

The reason for not finding any tokens of the word 'that' functioning as a determiner rather than a relative clause marker, I believe, is that two criteria have to be fulfilled for sentences of this kind to appear: a) intransitive verbs or verbs that have two objects (e.g., GIVE: give somebody something); and b) the indirect object of this verb should be expressed by a noun phrase beginning with that (e.g., that apple, that desk), which is not so commonly heard, because there are other variants (e.g., this, these, those) in this position. Thus for this part of the study, the colligation pattern N + is used.

Figures 4 and 5 show that the two kinds of colligational patterns are used in a similar way in terms of frequency and distribution in CEC and BNC, suggesting that they are largely similar in terms of colligational patterns in the expanding circle variety (CE) and inner circle variety (BrE) being examined.

4.6 Conclusion of colligation features of CE

The results of the analysis in this section suggest that CE users adhere to the basic norms of the inner circle variety in terms of colligation, but CE is also highly selective in the case of certain patterns. Being consistent with the British norm allows CE to be seen as acceptable, while the choice of some preferred patterns allows CE to maintain its identity and characteristics.

One instructive example of this is shown by the proportion of overused CE collocations (relative to their counterparts in BNC), where it becomes clear how central to CE these colligations are. These findings suggest that colligational differences make up a significant part of CE.

5. Implications and Conclusions

Based on the findings in this study, there is evidence showing that the modifier−modified sequence before the head noun is preferred by CE users, resulting in the salience of sometimes lengthy pre−nominal modifying structures. Word order is one of the most powerful devices used in Chinese to indicate subtle changes in meaning. This governing principle requires the word order of a modifier to appear before the noun it modifies in Chinese, no matter whether the modifier is an adjective, an attributive clause, a prepositional phrase, an infinitive verb phrase (e.g., 可愛的孩子, 媽媽買的衣服, 朝南的房子, 跳的高度) (Y. H. Liu, 2001, pp. 46−47). Most of the Chinese syllables are morphemic; any alteration to their sequence will lead to a significant change in their meaning.

For instance, purely through word order, the following *shunkouliur* (顺口溜) 'Chinese doggerel' in the following example suggests that the citizens of the provinces mentioned differ as to their tolerance of or predilection for spicy food (W. Jiang 2009: 4). The first line of the example is written in Chinese characters. The second line is the same sentence written in Pinyin, the official Chinese phonetic system used in the People's Republic of China. This is followed by a word−for−word or literal English translation in the third line. The last line in the example provides an idiomatic English translation.

(1) a 四川人不**怕辣**
Sìchuānrénbùpàlà;
Sichuan person not fear spicy;
Sichuaners do not fear (their food) being spicy.

(1) b 湖北人**辣**不**怕**
Húběirénlàbùpà;
Hubei person spicy not fear;
(Their food) being spicy is not a fearful matter to Hubeiners.

(1) c 湖南人**怕**不**辣**
Húnánrénpàbùlà.
Hunan person fear not spicy.
Hunaners fear that (their food) is not spicy (enough).

The very subtle differences in meaning in these sentences (1) *a* to (1) *c* are expressed by rearranging the word order of the last three morpho-syllables: *bù* (in blue) meaning 'not', *pà* (in green) meaning 'fear', and *là*(in red) 'spicy (food)'. The colligations of these three phrases are, respectively:

(1) a Negator + verb + noun (不**怕辣**, a collocation of *not fear* + *sth*)

(1) b Noun + negator + verb (**辣**不**怕**, topic-comment structure, with N functioning as topic, and the negated verb phrase serving as comment)

(1) c Verb + negator + adjective (**怕**不**辣**, V-O structure, in which the object position is filled by a negated adjective or stative verb '(being) spicy')

* 辣(*là*) may be used as a noun or an adjective in Mandarin, depending on the syntactic position or collocational pattern in which it appears. For example, in 很辣 (*hěnlà*, 'very spicy'), it serves as an adjective, while in 吃辣 (*chīlà*, 'eat spicy food'), it is a noun.

Generally speaking, the three phrases express very similar ideas; that is, Sichuaners, Hubeiners, and Hunaners are alike, in that they all like spicy food very much. Word plays a part, varying the word orders as a rhetorical device to convey subtle differences in meaning is not rare in Chinese. Word order (here in this study limited to noun phrases) is, therefore, an important aspect of the language disposition, influencing the structuring of information that CE users rely on heavily to express or manipulate meaning for special semantic effect. Rearrangement of the word order might not only change the meaning but could also shift the intended rhetorical effect of a given idea. For example, compare the effect of telling one's teacher 我不完全懂 (wǒ bù wán qúan dǒng, I did not understand fully [about what you said]) and 我完全不懂 (wǒ wán qúan bù dǒng, I did not catch anything [you said]). While the five morpho-syllables ('not completely V' vs. 'completely not V') are identical, varying their order would not only change the scope of negation and result in rather different meanings, but it would also result in rather different rhetorical effects. Thus in Mandarin, word order plays a crucial role; in that it helps people deliver information and ideas in subtle ways. To play safe, therefore, the CE user would be inclined to use the structure they feel more comfortable with, thus making the modifier-modified sequence in their noun phrases statistically such a distinctive feature in CE.

Thus the empirically supported answer to our research question 'Are there colligational preferences in CE?' is: Yes, there are distinctive colligational features in CE. One of the features of CE found is that users tend to demonstrate a clear preference for putting the modifier in front of the modified rather than after it, thus forming a modifier-before-head sequence when constructing collocations. More specifically, within a noun phrase, most

modifiers (adjectives in this study) occur before the head nouns.

Furthermore, our analysis of CE data yielded far fewer cases of the *N + of* structure compared with BrE, which indicates a post-modifying or reverses word order in the structure of elements within a noun phrase. So a more specific answer to a question like 'Are there colligational preferences in CE?' is that colligational preference is indeed found in CE, in particularly preferred word order patterns. According to Lian (2010: 25), Mandarin Chinese, like other Chinese 'dialects' or languages in the Sino-Tibetan family, is typologically an analytical language, with little inflectional morphology to convey grammatical relationships. An analytic language is marked by the relatively frequent use of function words, auxiliary verbs, and varying word order as principal means to express syntactic relations rather than relying on inflected word forms. As a result, free morphemes, which are often separate words, are used very commonly in grammatical constructions along with word order. In sum, one of the distinctive features of the L1-conditioned cognitive disposition of CE users is derived from the fact that Chinese languages and dialects have relatively restrictive word orders, often relying on the order of constituents to convey important grammatical information. In contrast to this, inner-circle varieties such as British English can convey grammatical information through inflection, which allows for more flexibility in terms of word order (e.g., both pre-modifying and post-modifying structures in a noun phrase). In the case of CE, the modifier-modified sequence is a word order structure that is used with a higher frequency than can be accounted for by chance, as reflected in the preference of CE users in the CEC corpus.

The author hopes to argue, by the above results, that the codifications of CE could move from an overt layer (e.g., collocations like *four modernization, xiaokang society*), to a covert layer — the colligational patterns that are hidden in English used by Chinese people.

Notes

1 The software used in this study is AntConc (Version 3.5.8), written by Anthony, L. (2019). Available from https://www.laurenceanthony.net/software.

2 The corpus used in this study, *China English Corpus(CEC)*, is provided by Professor Li Wenzhong through personal contact in 2011. The author would like to express her sincere gratitude to Professor Li and his team for their kindness and hard work on the project of CEC.

3 The author would like to thank Professor David C. S. LI for his advice on this paper.

References

Adamson, B., Bolton, K., Lam, A. & Tong, Q. S. (2002). English in China: A preliminary bibliography. *World Englishes*, 2, 349-355.

Adamson, B. (2004). *China's English: A history of English in Chinese education*. Hong Kong: Hong Kong University Press.

Adamson, B., & Feng, A. (2014). Models for trilingual education in the People's Republic of China. In Gorter, D., Zenotz, V. & Cenoz, J. (Eds.), *Minority languages and multilingual education: Bridging the local and the global*. Netherlands: Springer.

Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

Bolton, K. (2003). *Chinese Englishes: A sociolinguistic history*. Cambridge: Cambridge University Press.

Bolton, K. & Graddol, D. (2012). English in China today. *English Today*, 1, 3-9.

Bolton, K. & Botha, W. (2015). Researching English in contemporary China. *World Englishes*, 2, 169-174.

Bolton, K., Botha, W. & Zhang, W. (2015). English in China: A contemporary bibliography. *World Englishes*, 2, 282-292.

Chen, X. (2010). Discourse-grammatical features in L2 Speech: A corpus-based contrastive study of Chinese advanced learners and native speakers of English. Ph.D. dissertation, City University of Hong Kong.

Cheng, C. C. (1992). Chinese varieties of English. In Kachru, B. B. (Ed.), *The other tongue: English across cultures*. Urbana: University of Illinois Press.

Crystal, D. (2001). *Language and the Internet.* Cambridge: Cambridge University Press.

Crystal, D. (2012). A global language. In Seargeant, P. & Swann, J. (Eds.), *English in the world: History, diversity, change*. London: Routledge.

Evans, S. (2011). Hong Kong English and the professional world. *World Englishes*, 3, 293-316.

Gao, C. (2008). A Corpus-based study of the use of creation and transformation verbs in China's English newspapers. Ph.D. dissertation, Nanjing University, Nanjing, China.

Ge, C. G. (1980). Random thoughts on some problems in Chinese-English translation. *Chinese Translator's Journal*, 2, 1-8.

Gui, S. (1988). *Applied English and English language teaching in China.* Ji' nan: Shandong Education Press.

Haberland, H. (2013). ELF and the bigger picture. *Journal of English as a Lingua Franca*, 1, 195-198.

Haberland, H., & Mortensen, J. (2012). Language variety, language hierarchy, and language choice in the international university. *International Journal of the Sociology of Language*, 216, 1-6.

He, D. Y., & Li, C.S. (2009). Language attitudes and linguistic features in the "China English" debate. *World Englishes*, 1, 70-89.

Ho, Y. (1993). *Aspects of discourse structure in Mandarin Chinese*. New York: Mellen University Press.

Hu, X. Q. (2004). Why China English should stand alongside British, American, and the other "world Englishes". *English Today*, 2, 26-33.

Hu, X. Q. (2005). China English, at home and in the world. *English Today*, 3, 27-38.

Hu, W. (1995). Functional perspectives and Chinese word order. Ph.D. dissertation, Ohio State University.

Huang, J. Q. (1988). The positive role of Sinicism in the English-translated version. *Chinese Translator's Journal,* 1, 39-47.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.

Jiang, Y. J., & Du, R. Q. (2003). Issues on "China English". *Foreign Language Education,* 24, 27-35.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching.* Cambridge: Cambridge University Press.

Kirkpatrick, A. & Xu, Z. C. (2002). Chinese pragmatic norms and "China English". *World Englishes*, 2, 269-280.

Leech, G. N. (2002). The importance of reference corpora. Keynote speech at *Hizkuntza-corpusak. Oraina eta geroa.*

Leech, G. (2007). New resources, or just better old ones? The holy grail of representativeness. *Language & Computers*, 1, 133-149.

Leech, G., Deuchar, N. & Hoogenraad, R. (2006). *English grammar for today: A new introduction*. Basingstoke: Palgrave Macmillan.

Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Li, C. N. & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, C. N. & David, C. S. (2002). Pragmatic dissonance: The ecstasy and agony of speaking like a native speaker of English. In Li, D. C. S. (Ed.), *Discourses in search of members*. Lanham: University Press of America.

Li, W. Z. (2005). *Extracting the multiword units from China's English news articles.* Paper presented at the BAAL corpus linguistics SIG/OTA workshop: Identifying and researching multiword units. Oxford University.

Lian, S. N. (2010). *Contrastive studies of English and Chinese.* Beijing: Higher Education Press.

Mackey, W. F. (1967). *Bilingualism as a world problem.* Montreal: Harvest House.

McArthur, T. (2003). World English, Euro-English, Nordic English? *English Today*, 1, 54-58.

McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory, and practice*. Cambridge: Cambridge University Press.

Mesthrie, R. & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Niu, Q. & Wolff, M. (2003). China and Chinese, or Chingland and Chinglish. *English Today*, 4, 30–35.

Pan, L. & Seargeant, P. (2012). Is English a threat to Chinese language and culture? *English Today*, 3, 60–66.

Sinclair, J. (2004). *Trust the text: Language, corpus, and discourse*. London: Routledge.

Sun, L. (2011). Degree adverbs in Hong Kong and Singapore English: A corpus-based investigation. Ph.D. Dissertation, The University of Hong Kong.

Tai, J. H-Y. (1985). Temporal sequences and Chinese word order. In Haiman, J. (Ed.), *Iconicity in syntax*. Amsterdam: John Benjamins.

Wang, W. B. & Ma, D. (2002). Chinese English and its expressions. *Journal of Dalian Nationalities University,* 14, 55–8.

Xu, Z. C. (2010). *Chinese English: Features and implications*. Hong Kong: Open University of Hong Kong Press.

Yu, X. (2009). Nativized lexical use in China's English newspapers. Ph.D. dissertation, Nanjing University, China.

Zhuang, Y. C. (2000). Guard against Chinglish. *Chinese Translators Journal*, 6, 7–10.

Zipf, G. K. (1935). *The psycho-biology of language*. Boston: Houghton-Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Boston: Addison-Wesley Press.