The Corpora of China English: Implications for an EFL Dictionary for Chinese Learners of English

Lixin Xia, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, Guangzhou, P.R. China (200211103@oamail.gdufs.edu.cn)*

Yun Xia, *Department of International Studies, Nanfang College of Sun Yat-Sen University, Guangzhou, P.R. China (344192304@qq.com)*

Yihua Zhang, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, Guangzhou, P.R. China (bilex@mail.gdufs.edu.cn)*

and

Hilary Nesi, *School of Arts and Humanities, Coventry University, Coventry, United Kingdom (aa3861@coventry.ac.uk)*

Abstract: The localization of the English language in China has brought about a distinctive English variety which has come to be known as China English. Recently, several corpora of China English have been or are being built; these will help us to identify the established linguistic features of this variety, and should greatly facilitate the compilation of an English dictionary for Chinese learners of English who operate in Chinese cultural contexts and need to refer to China-specific concepts and phenomena. This paper briefly introduces China English corpora in terms of their principles, components and current status, and explores their potential application to lexicographical projects in terms of lemma selection and inclusion, definition extraction, and glossing and labeling systems.

Keywords: CHINA ENGLISH, CORPUS, DEFINITIONS, ENGLISH–CHINESE LEARNER'S DICTIONARY, ENGLISH VARIETIES, GLOSSES, HEADWORD SELECTION, LABELS, LEXICOGRAPHY, LINGUISTIC FEATURES OF CHINA ENGLISH, LANGUAGE CONTACT, WORLD ENGLISHES

Opsomming: Die korpora van China-Engels: Implikasies vir 'n EVT-woordeboek vir Chinese aanleerders van Engels. Die lokalisering van die Engelse taal in China het 'n eiesoortige Engelse variëteit tot stand gebring wat as China-Engels bekend geword het. Onlangs is of word verskeie korpora van China-Engels saamgestel; dit sal ons help om die gevestigde taalkundige kenmerke van hierdie variëteit te identifiseer, en dit behoort die samestelling van 'n Engelse woordeboek vir Chinese aanleerders van Engels wat binne die Chinese kulturele kontekste optree en wat moet verwys na konsepte en verskynsels spesifiek tot China grootliks te vergemaklik. Hierdie artikel stel kortliks China-Engelse korpora bekend volgens hul beginsels, komponente en huidige status, en ondersoek hul moontlike toepassing op leksikografiese projekte

rakende lemmaseleksie en -insluiting, die onttrekking van definisies en stelsels vir glossering en etikettering.

Sleutelwoorde: CHINA-ENGELS, KORPUS, DEFINISIES, ENGELS–CHINESE AANLEER-DERSWOORDEBOEK, ENGELSE VARIËTEITE, GLOSSE, LEMMASELEKSIE, ETIKETTE, LEKSI-KOGRAFIE, TAALKUNDIGE KENMERKE VAN CHINA ENGELS, TAALKONTAK, VARIË-TEITE VAN WÊRELD-ENGELS

1.   Introduction

Ge (1980) seems to have been responsible for coining the term "China English". He did not define it precisely, and initially scholars debated whether it existed as a phenomenon distinct from 'Chinese English' or 'Chinglish'. Even now, China English has not been universally acknowledged as a variety, but it is becoming a topic of growing interest to scholars in China and internationally, and it is generally believed to present its own phonological, lexical, syntactic, pragmatic, and discourse features (Cannon 1988, Jiang 2002, Du and Jiang 2001, Xu 2008, Xu 2010b, Bolton 2006, Kirkpatrick 2010, Yu and Wen 2010).

Ge (1980) gave examples of China-specific phenomena and concepts which he believed to be representative of China English, such as *Four Books* (*Si Shu*), *May Fourth Movement* (*Wusi Yundong*), *ideological remoulding* (*sixiang gaizao*), and *four modernizations* (*sige xiandaihua*). His ideas were expanded by Li (1993: 19) who went on to describe China English as a variety of English which has Standard English at its core, but is used to express China-specific referents and concepts in the context of Chinese culture. The extent to which it is influenced by the Chinese language remained unclear.

According to Xu (2006, 2008, 2010a, 2010b), China English should be defined as:

> a developing variety of English, which is subject to ongoing codification and normalization processes. It is based largely on the two major varieties of English, namely British and American English. It is characterized by the transfer of Chinese linguistic and cultural norms at varying levels of language, and it is used primarily by Chinese for intra- and international communication. (Xu 2010a: 1)

At the phonological level, some patterns in China English are predictable, and are handed down from one generation to another. For example, Chinese users are confused by the stress rule in English (Pan 2005: 49) as Chinese is a tone language. Jiang (2002: 12) notes that for the verb + adverbial pattern, the stress is generally on the verb, so speakers of China English say *SIT down* and *STAND up*, rather than *sit DOWN* or *stand UP*. To an English native speaker the former might sound more like an order, and they would generally use the latter forms except when emphasizing the command. He and Li (2009: 72) report the following phonological features of China English: replacement of /θ/ with /s/

and /ð/ with /z/, insertion of final [ə], avoidance of weak forms for function words, and general lack of voiced fricatives.

It is common for lexical items to enter the vocabulary of Standard English from Chinese; for example Chinese expressions accounted for 20% of the 2,000 new words and phrases added to English in 2005, according to the Global Language Monitor (Radtke 2007). Most of these are accepted into English because of the need to express key features of the physical and social environment in China peculiar to Chinese culture. However we can imagine that such expressions are established within China English some time before becoming fully accepted into Standard English.

The following non-standard forms, identified by Kirkpatrick (2010) in the Asian Corpus of English, are widely attested in China English.

1.   The non-marking of the third person singular with "-s"
2.   Interchangeability of the relative pronouns, "who" and "which"
3.   Flexible use of definite and indefinite articles
4.   Extended use of "general" or common verbs
5.   Treating uncountable nouns as plural
6.   Use of a uniform question tag
7.   Use of demonstrative "this" with both singular and plural nouns
8.   Use of prepositions in different contexts (Kirkpatrick 2010: 8)

Xu (2010a: 69) identifies additional syntactic features which are typical of China English:

1.   Adjacent default tense (So, <u>I have to sign</u> a contract <u>before I came here</u> for the study.)
2.   Null-subject/object utterances (A: When you're at home, you talk in Fujian dialect with your parents? B: Sometimes yes, sometimes__use Putonghua.)
3.   Co-occurrence of connective pairs (<u>Although</u> it is not as big as Beijing, <u>but</u> I like it.)
4.   Subject pronoun copying (<u>Some of my college classmates they</u> like to dress up very much.)
5.   Yes–no responses (A: You do not want to make a living by playing guitar on the street. B: <u>Yes.</u> Of course not.)
6.   Topic–comment (I think <u>this society</u>, the people get more and more practical.)
7.   Unmarked OSV (Probably <u>some other kind of jobs</u> <u>I</u> also <u>want to try</u>.)
8.   Inversion in subordinate finite wh-clauses (I really don't know <u>what is</u> international English.)

But some of Xu's features listed above are at the discourse level, such as yes-no response and co-occurrence of connective pairs.

At the discourse and pragmatic level, Chinese characteristics occur in China English in almost every aspect. According to Kirkpatrick (2002), the

frame-main sequence in Chinese is expected to be transferred into China English. As one might expect, China English is greatly influenced by sociopragmatic considerations relating to political status and political life, law and social order, power and hierarchy (Xu, 2010 a). He also reports some pragmatic features of China English, for example, through the analysis of culturally unique expressions in written texts.

2.   The Teaching of China English in China

English is widely taught in China from primary schools to universities and colleges; a national language survey conducted at the turn of the century found that 390.16 million Chinese people had learnt English at some stage in their education (Wei and Su 2012).

Currently the pedagogical model for English in China adheres to native speaker norms, although this is an unattainable goal because, as Medgyes (1992: 342) puts it, "non-native speakers can never achieve a native speaker's competence". Moreover, in the age of globalization the goal should be "to use English successfully in lingua franca or multilingual contexts, rather than to acquire standard forms and nativelike proficiency (Kirkpatrick 2011: 221)".

It is for this reason that researchers and educators such as Hu (2007), Pan (2005), Wu (2014) and He and Li (2009) argue that China English should be taught at schools alongside British or American English. The obvious advantage for this new pedagogical model of English is that the inclusion of selected linguistic and pragmatic features of China English will provide learners with more opportunities to express ideas specific to Chinese culture. The latent advantage is that it helps learners of English in China shake off the fetters of 'Standard English', and develop communicative competence in the context of cross-cultural communication between users of different Englishes.

3.   Corpora of China English

China English corpora are usually made up of texts written or translated by Chinese authors. The following sections provide a brief review of each of the major corpora of this type.

3.1   The China English Corpus

In 2009, Li Wenzhong initiated the construction of the China English Corpus (Sun and Lv 2010). This corpus consists of written texts, mostly taken from newspapers, magazines, academic journals, dissertations, literary writings, government documents and web pages. To keep the original form of the data, non-textual data such as photos and graphs were also input into the corpus

through links provided in the annotation so the users can retrieve and extract original texts. The data is given linguistic tags.

Before building the China English Corpus, a pilot project was launched to build a corpus of China English News Articles (CENAC). This pilot corpus contained 2,316 texts which are classified into twelve categories: life, politics, economy, law, environment, education, science, sports, communication, advertisement, agriculture, and industry. All the texts came from news articles published in the *China Daily*, *Beijing Weekend*, and *Shanghai Star* in 2002. It has 1,281,498 tokens.

According to Li Wenzhong, the China English Corpus was built by the corpus research and development team of Henan Normal University (Li 2012: 336). However, literature on it is extremely rare except for some papers introducing the initial annotation system and using CENAC to study the word clusters in news articles (Li 2007, Liu 2008, Sun and Lv 2010, Gao 2006). Moreover, the size and status of this China English Corpus are unknown. It is not accessible for public use. Fang (2012: 113) explains that "the corpus is still under construction."

3.2   *China Daily* Corpus of China English

The China Daily Corpus was built by Li Dingshun in 2011 (Li and Liu 2011). All the data came from the *China Daily* and *China Daily Hong Kong Edition*. It has a total of 156,864 texts, 120 million tokens, and 350,000 words. The corpus covers the following 10 domains: politics, economy, culture, life, sports, education, entertainment, environment, technology, and military.

This corpus has been completed, but it is only representative of the *China Daily*; besides, the Hong Kong component of the corpus will represent a rather different cultural environment from that of mainland China.

3.3   The Comparable Corpus of China English

This corpus is being built by Fang Chengyu (Fang et al. 2012) in Hong Kong for the purposes of comparing China English and British English. The corpus is made up of three modules: media English, academic English, and personal microblogs. At present, the first module is complete, and enables comparison of Chinese and British English as used in newspapers, magazines, and web pages, across the following domains: news reports, editorials, life, culture, and business.

The Chinese Media English Corpus has a total of 2,226 texts. Each text has been cut so that only the first 600 words have been included, so the sub-corpus contains 1,125,627 tokens. The British Media English Corpus was compiled in the same way as the Chinese one and has 1,133,643 tokens. The second module representing academic English is currently being built.

This comparable corpus is particularly useful for comparison of the linguistic features of Chinese and British English. By including both Chinese and British English in the corpus, it should reveal aspects of China English that the other corpora cannot. However, the size of the corpus is modest compared with other modern corpora.

3.4   China-Related English Corpus

This is a project led by Zhang Yihua at Guangdong University of Foreign Studies supported by the National Social Science Fund of China. The project began in 2013.

The corpus is made up of two sub-corpora: China-related English by Chinese native speakers and China-related English by English native speakers. For the first of these, data was collected from English media sources in China, such as the *China Daily*, *Beijing Review*, *China Today*, *Global Times*, and from novels and stories written in English by Chinese people or translated from Chinese into English by Chinese people. For the second sub-corpus, data was collected from news and reports about China and Chinese culture in British and American media, such as the *BBC*, *Reuters*, *VOA*, *CNN*, *ABC* and the *New York Times*, and from books and stories on China and Chinese culture written and published in the UK and the USA, and from works written by Chinese but translated into English and published in the UK and the USA.

At present, more than 150-million tokens have been collected and processed, and a pilot version is available for public use (http://202.116.192.97/#). The corpus is scheduled to be completed by 2017.

This corpus is by far the largest and most comprehensive one of this type in terms of its size and genres it covers. It may be the only one of this type that is designed and constructed with lexicography as its primary aim. To the designers of this corpus, China-related English means English on China and Chinese culture no matter whether it is written by Chinese people or English native speakers. From the components of the corpus, it can be inferred that it enables comparison of China English and 'standard' British or American English at different levels. However, one may argue that it is difficult to identify 'English native speakers'. Just because they write in English in Britain or the USA doesn't necessarily mean they are native speakers of English.

All the corpora mentioned in this section are made up of written texts only; and none of them contains spoken data. It is generally agreed that varietal differences are much more evident in speech than in writing. Therefore, corpora of this type should be made up of both written and spoken data. Moreover, the issue of representativeness should be tackled carefully as China is so big.

4.   Corpora of China English and Lexicography

Dictionary production can be seen as a four-stage process: deciding the target

user, collecting linguistic data, analyzing linguistic data, and synthesizing dictionary text (Rundell 2010: 367). At the first stage, it is important to note that dictionary users may be composed of speakers of different Englishes in cross-cultural communication. The globalization of English has caused "lexicographers to think about the adequacy of their products for a diverse and complex world of English-users (Benson 1997: 126)". To meet the needs of the users, an English dictionary should record and describe the actual use of world Englishes rather than solely that of British or American English.

As a result, some regional English dictionaries or dictionaries of a local English variety have been made. *The Macquarie Dictionary* (1981) was the first comprehensive dictionary of Australian English (Delbridge and Butler 1999), and the most innovative aspect of its third edition (1997) is that it includes words from Southeast Asian English "which have never been in a dictionary before" (Butler 1997: 103). The *Dictionary of Caribbean English Usage* (1996) employs a specially-designed labelling system to denote the social or grammatical register of Caribbean English (Allsopp 1996). *Cosmo Dictionary of Indian English* (Yule et al. 2006) provides a ready reference for words and phrases of Indian English not included in other English dictionaries. *A Dictionary of South African Indian English* (Mesthrie 2010) records the use of Indian English in South Africa, and it contains around 1700 entries. These dictionaries are all innovative because they included non-western varieties of English.

It can be predicted that more and more national English dictionaries of local varieties will be joining *the Macquarie Dictionary* in the foreseeable future. It might not be the right time to compile a dictionary of China English. However, the EFL dictionary specifically designed for Chinese learners of English should take the lead in including and describing words and use of China English in order to meet their needs in expressing concepts and phenomena specific to China and Chinese culture. Furthermore, the dictionary should provide information on linguistic and cultural features of China English systematically in its definitions, glosses, labels, etc.

4.1   Criteria of Lemma Selection and Inclusion

"There is general consensus among scholars and researchers of 'China English' that the defining feature of 'China English' is its lexicon (He and Li 2009: 72)." That being the case, it would be natural for a dictionary user to expect a substantial amount of Chinese words in an English dictionary. However, it is not the case that there has been substantial, or even adequate inclusion and description of them. In the process of writing a dictionary, any item from an English variety in Kachru's expanded circle (1985), like China English, has to fight hard in order to enter the English dictionary. One may argue that it is because of their low frequency in a corpus that they are not given the right to enter a dictionary as a headword since all major English dictionaries are based on corpus data. The inclusion criteria themselves seem to be scientific. But if

the corpora on which these English dictionaries are made do not include texts from China, problems arise. The outcome would be the loss of certain words of Chinese origin from dictionaries. For example, the author extracted some of the texts from the first sub-corpus of the China-related English Corpus for the purpose of this study. It has about 2 million tokens, and will be used as an acting corpus of China English (hereafter ACCE) for this study. A search of the English word *gaokao* from Chinese in the ACCE got 57 concordance hits. But there are no hits of the word in the BNC which has 112,289,776 tokens (data searched on 4 July, 2016).

These items are also excluded because of long-standing lexicographic traditions. In the fields of linguistics and lexicography, British English and American English are looked up to as the standard. But many varieties of English especially those from the expanded circle have not been accorded the status of a legitimate variety of English. They are regarded as non-standard Englishes. Although contemporary lexicography aims to describe not prescribe, a dictionary has an inherent nature of standardization to the extent that once any of a linguistic item is included in a dictionary, the dictionary raises "it to the status of the norm (Svensén 2009: 67)". Therefore, it's safer to exclude these 'non-standard' items from a dictionary.

Although the mainstream English learner's dictionaries such as the 'Big Five' claim to be specifically designed for foreign learners, few lemmas from Chinese are included in them. There are only 13 words borrowed from Chinese in the *Oxford Advanced Learner's Dictionary* (8th edition, 2010, CD-ROM version) (*cheongsam, mahjong, samfu, taipan, wok, wonton, yang, yin, fengshui, dim sum, kung fu, foo yong,* and *t'ai chi ch'uan*), although it claims to include as many as 184,500 headwords. That is to say, words of Chinese origin only account for 0.007% of all the headwords. Benson (1997: 133) notes that these learners' dictionaries not only contain fewer references to China than their larger counterparts such as the OED, but also that these references form a much smaller proportion of the total number of entries — 0.12% for OED 2, and 0.08 for OALD4. The OED online version (*http://www.oed.com*) now has a total number of 600,000 entries. Through the advanced search, one can get 251 entries of Chinese origin (data retrieved on 3 July, 2016), which makes up 0.041% of the total entries. Nearly 20 years passed from Benson's work, but the situation remains unchanged.

Dictionary writers are extremely cautious about admitting new entries to their dictionaries. They have more reasons not to include an alien word from Chinese since it is in the expanded circle of the English family (Kachru 1985). However, Chinese English learners are brought up in the context of China, and they have to express something typical of China and Chinese culture. If a dictionary doesn't include them, they will not know how to express these culture-specific ideas. That may account for the incompetence of Chinese learners when they want to express some specific Chinese referents and concepts.

It therefore seems a good idea to create an EFL dictionary for Chinese learners of English, including more entries relating to China-specific concepts. A search in the China-Related English Corpus can generate hundreds and

thousands of hits for words and expressions with Chinese characteristics. We find that most of the single words are in their Pinyin form, such as *laojiao* and *gaokao*. However the multi-word expressions are mostly in their anglicized form, such as *educated youth*, *four virtues*. Since these expressions are heavily culture-bound, it is almost impossible to guess their meanings from the meaning of their components. Therefore, both types of lemmas need to be included and explained in the EFL dictionary for Chinese users. In summary, the criteria of inclusion should take into account the corpus evidence from corpora of China English for a word's frequency and use.

4.2   Defining Styles

The major contribution of corpora of China English to lexicography is that they enable us to reexamine the ways in which lemmas from Chinese are defined in English dictionaries. In general, these lemmas are not accurately defined in them. One can even say that they are often vague, and sometimes even misleading. For example, a native speaker of Chinese may fail to figure out what the headwords refer to in the real world by reading the definitions of the following two entries: *cheongsam,* and *samfu* in OALD8. The lemma *cheongsam* is defined as "a straight, tightly fitting silk dress with a high neck and short sleeves and an opening at the bottom on each side, worn by women from China and Indonesia". Native speakers of Mandarin Chinese have a different picture of the cheongsam, as shown in Figure 1. In mainland China it is understood to be a long and loose-fitting piece of clothing which covers the body and reaches the ground; formerly in China cheongsams were especially worn by educated men, and were a sign of rank because poor, illiterate people couldn't afford them.

Figure 1: An illustration of *cheongsam*

The definition in OALD8 is completely different, and refers to an item of clothing that is known as a *qipao* in Mandarin Chinese (see Figure 2), and is translated as such in the bilingualised version of OALD8. The translator might have spotted the mismatch between the headword and the definition, but the error has not been corrected.

Figure 2: An illustration of *qipao*

The word form *cheongsam* is Cantonese, and the dress originated from Shanghai in the 1920s, as an exclusively traditional gown for women. For Cantonese speakers in the areas of Guangzhou, the word *cheongsam* usually constitutes a jacket with long sleeves, not necessarily a long dress covering the whole body. In Hong Kong, the word *cheongsam* refers to a dress for both women and men. It seems that OALD8 adopted the signifier of the concept in Cantonese and the signified object from Shanghai areas. No regional uses were shown in the definition (Xia and Zhai 2016). That's the reason why speakers of Mandarin Chinese would feel puzzled by the definition in OALD8.

The headword *samfu* is defined as "a light suit consisting of a jacket with a high collar, and loose trousers/pants, worn by Chinese women" in OALD8. A speaker of Mandarin Chinese might not be able to figure out what a *samfu* is by reading the definition. Unfortunately, this entry has not been included in the bilingualised version of OALD8. Maybe the translator found the problem, and dropped the entry from the original dictionary deliberately.

As mentioned above, these words and expressions from Chinese only account for an extremely small proportion of the entries in the English dictionary. So it is not a big problem for Chinese English learners even if they are not defined accurately. They have bigger problems with common English words and expressions. For example, the first sense of the lemma *house* in OALD8 reads:

> **house¹** *noun* … **1** [C] a building for people to live in, usually for one family:
>   The family live in a two-storey house… .

In Britain and America, a house is often a detached, semi-detached or terraced building with two or three levels for one family to live in. The definition is concise and good for them, but not for Chinese English learners. Here is a concordance line from ACCE of China-related English Corpus:

> House prices in Beijing are absolutely ridiculous. When a young couple purchases a <u>house</u>, parents and grandparents from both sides need to help out. The collective effort of four families is required to support a young family's decision to buy a <u>house</u>. Three generations of savings are thus exhausted in buying a single <u>house</u>.

House prices are too high for a young couple to buy even a small home in cities in China. Their parents and grandparents have to help them buy one with their savings. They are buying an apartment in a high building for the young couple, not a two-storey house. They can't afford an apartment, let alone a big house. Therefore a house in China English as shown in the above example is a set of rooms in a high building for people to live in or for rent, something like an apartment or a flat as shown in Fig. 3.

Figure 3: An illustration of *house* for people in the city

For older Chinese people, especially those from the countryside, a house is a small building with a tiled roof, on one level as described below in the concordance line from ACCE, although more and more people in the rural areas are building multi-storey houses now.

> Traditionally, a typical rural residence is a three-room, one-storey house with a spacious courtyard where trees and flowers can be planted.

Figure 4: An illustration of *house* for older people in the countryside

The closest equivalent to a house is probably *bieshu* as shown in Fig. 5. A bieshu in China refers to a detached or semidetached building with two or three levels for one family, especially a rich one.

Figure 5: An illustration of *bieshu*

More concordance lines of 'house' can be found in ACCE, some of which are listed below:

1. a bird's-eye view of past and future: squat tile-roofed **houses** built around courtyards
2. ain. Now he has bought a **house** for the couple and hopes that one day they can wed.
3. My wife and I bought a modest **house** in suburban Shanghai a few years ago, and on
4. in the capital. "I'm getting ready to buy a **house** in Beijing, a big house," Mo said,
5. prospective husbands are vetted about whether they own a **house**, and preferably also
6. is already pregnant again and he is building a new **house** on the site of his old house
7. ts who hover over their children, "doing their work for them and **buying** them houses.
8. is very poor, his house is very shabby, but he always does good things for the villager

9.   We can't afford to rent an apartment if we want to buy our own **house**," Liu said. From

10.   takes a reporter to see the former village leader's home — a three-story **house** whose

11.   Wang Hong sat inside a small stone-and-brick **house** with one of China's aging musical

12.   in front of his traditional brick house a few yards away, Kong Qingyu, 77, a farmer, spo

13.   He revisited his clan village and his **house**, now abandoned, and met an elderly man

From the contexts it may be inferred that the word 'house' in concordance lines 1, 6, 8, 10, 11, 12 and 13 refers to a building in the countryside. Most of the houses are single-storey buildings except in line 10 where the reference is to a three-storey house. The owner of the house is the village leader. However, the word 'house' in concordance lines 2, 3, 4, 5, 7 and 9 probably refers to an apartment or flat in a high building in the city, not a two or three-storey house. Although it is not clearly indicated in the sentences, one can infer it through the contexts and background knowledge. As house prices are ridiculously high in China, most citizens cannot afford a house, especially in big cities like Shanghai (concordance line 3) and Beijing (concordance line 4). Most urban residents live in an apartment or a flat.

From the above analysis, we can see that the definition of the English word *house* will probably form different images in the mind of Chinese users. Therefore, to help Chinese learners better understand the meaning and use of the English lemmas, they should be defined in a way that accords with Chinese learners' cognitive schema. The most important thing is to tell the dictionary users the difference between China English and British English. So the definition of *house* could be like this:

> **house** *noun* [C] **1** a building for people to live in, usually for one family **SYN. villa 2** In China a house is a small building that has a tiled roof and only one level in the countryside, or a set of rooms in a high building for people to live in or to rent in the town **SYN. apartment, flat, bungalow, home …**

Compared with the definition in OALD8, a new sense of the headword *house* is given, depicting clearly the referent in the real world in China English. In other words, the English word *house* in China English most probably denotes two kinds of buildings: a single-storey building and a multi-storey high building.

By reading the definitions, Chinese English learners would recognize the difference between the extensions of the word house in 'standard' English and China English. The word *house* in China English has a wider extension than that in 'standard' English. Thus, when they use the word house in different regional contexts, they would note its regional use. For example, when they mention a house in China, they could use a specific word such as *apartment*, *flat*, *home*, *villa*, *bungalow*, etc. to denote it instead of a more general word *house*.

The inclusion and definition of English lemmas with Chinese characteristics will undoubtedly enrich the English language since more English speakers will become familiar with them and use them in their communications.

4.3   The Glosses

Apart from lemma inclusion and definitions, glosses can be used to remind dictionary users of the characteristics of China English. China English has its particular linguistic features. For various reasons, some headwords have more than one spelling. However, of the 13 headwords of Chinese origin, OALD8 only gives a variant for *t'ai chi ch'uan* as below.

> **t'ai chi ch'uan ... (**also **t'ai chi)** *noun* [U] (*from Chinese*) a Chinese system of
> exercises consisting of sets of very slow controlled movements …

T'ai chi is often used as the shortened form of *t'ai chi ch'uan* both in the West and in China (Xia and Zhai 2016). However, this gloss is problematic since the word *t'ai chi* is also a philosophical term in ancient China, meaning the quintessence of the universe, or the Supreme Ultimate. The word *t'ai chi ch'uan* refers to a Chinese martial art and a system of calisthenics.

Data from corpora of China English also shows that the shortened form is used almost exclusively to refer to the sports activities. No concordance lines of *t'ai chi* are found to refer to the philosophical thinking. The reason for this might be that the sports activity is more popular among people than the philosophical thinking.

1.   after participating in 12 weeks of **t'ai chi** training. The study's lead a
2.   esized that the dancelike style of **t'ai chi** might be an appropriate way
3.   for 12 weeks for those learning **t'ai chi**. A control group, whose mem
4.   y, those who participated in the **t'ai chi** exercises reported improvem
5.   e attractive to them." A Beijing **t'ai chi** master agreed to conduct the
6.   than half of them continued the **t'ai chi** class after the study was over
7.   year-old mother-in-law studied **t'ai chi** in China for two weeks and
8.   raditional Chinese calligraphy, **t'ai chi**, and martial art to better und
9.   classes were taught by several **t'ai chi** masters included Zeng Haixia
10.   where many still practise their **t'ai chi** in the misty early mornings, w
11.   n joggers pant by still, and the **t'ai chi** practitioners silently exercise a
12.   groups of pensioners practising **t'ai chi**, Chinese children can be seen p
13.   likely candidate to be practising **t'ai chi** in a glass-fronted shop in Man
14.   dicine. She also recommended **t'ai chi** as a means of relaxation. "Every

In this situation, it is advisable to add a note to the entry in an EFL dictionary for Chinese learners of English, telling dictionary users the original meanings of the variant *t'ai chi*. The entry could be rewritten as below:

> **t'ai chi ch'uan ...** (also **t'ai chi)** *noun* [U] (*from Chinese*) a Chinese system of
> exercises consisting of sets of very slow controlled movements …
> **Note:** The word form *t'ai chi* as a philosophical term means the quintes-
> sence of the universe, but it is used more frequently as an informal form of
> *t'ai chi ch'uan.*

With the help of a corpus of China English, lexicographers can get all the vari-
ants of a lemma and their frequency. For example, through data retrieval in
ACCE, we get all the spellings of *kung fu (961), kungfu (99), kongfu (8),* and *gong
fu (8)*. Based on the word frequency, *kung fu* should be listed as the headword,
kungfu as the variant, by way of a gloss as shown below:

> **kung fu ...** (also **kungfu)** *noun* [U] (*from Chinese*) a Chinese sport people do to
> fight and stay healthy. Kung fu is a martial arts. …

In the above entry, *kungfu* is listed as a variant of the headword *kung fu* since it
is less frequently used, and the word *also* in the gloss indicates this. The word
forms of *kongfu* and *gong fu* are not listed as a variant because of low frequency.

At the grammatical level, typical characteristics of China English can be
indicated by glosses. Because of negative transfer from their first language,
speakers of China English would naturally use the collocation structure of their
mother tongue instead of a 'correct' English one. Yu and Wen (2010) made a
comparative study of the collocation patterns of evaluative adjectives and
nouns in China English and British English with data from CENAC (China
English News Articles Corpus) and the news part of BNC. They found that the
collocation patterns of adjectives and nouns in China English are greatly influ-
enced by Chinese. For example, the word *great* co-occurs with *change(s)*, *pro-
gress*, *contribution* and *achievement(s)* more frequently in China English than in
British English with a standardized frequency of 24:0.5, 11:0.5, 10:0.4 and 10:1
respectively. They argued that this collocation pattern was transferred directly
from Chinese. Therefore, this collocation pattern should be treated and repre-
sented in an EFL dictionary for Chinese learners of English.

Semantic prosody also needs to be considered in an EFL dictionary. Sin-
clair (1991: 121) noted that "many uses of words and phrases show a tendency
to occur in a certain semantic environment. For example the word *happen* is
associated with unpleasant things — accidents and the like". A search of *happen*
in BNC found that the following words co-occur with it: *accident*, *incident*,
attack, *crash*. However, during the use of English in China over the years, it has
acquired a unique flavor of its own. As a result, some words and phrases have
a different semantic preference from their original use in 'standard' English. For
example, the word 'happen' was found to be together frequently with positive
words in ACCE. Here are some examples:

1.   Though we are seeing signs that a **pick-up** might happen in the US economy, it
      certainly seems that China is not going to return to the heady days of 20% export
      growth a year.

2. You know, if **lots of lucky things** happen, we can get out relatively painlessly. But I think the probability of that is very, very low.
3. There needs to be **wide-spread application of these best practices** across the board in all new building and construction projects. I am looking forward to seeing that happen.
4. But over time, and with the help of business associations that helped firms develop their own capabilities, these firms evolved, became more sophisticated, more profitable, invested in modern production and human resources management techniques ... and eventually, all of these abuses went away. I expect **the same** can happen in China.
5. Frank Sha, a senior consultant with Zou Marketing, a sports branding company based in Shanghai, says that demand is strong among Chinese consumers for big events like Saturday's soccer match. What is lacking is the people to make **such events** happen regularly.
6. It hopes to launch the rover-releasing moon probe in about two years. Chinese experts believe **a moon landing** will happen in 2025 at the earliest.

The words in boldface in the above examples are generally associated with pleasant things, such as a pick-up in economy, lucky things, big events and a moon landing. It, therefore, can be concluded that the word *happen* tend to co-occur with words with both positive and negative associations in China English.

4.4   Labelling System

Due to the cultural differences between Chinese and Western culture, the same concept may have different connotations. These differences can be indicated in the EFL dictionary by means of a label or a usage note. But labels used for the English lemmas of Chinese origin in OALD8 tend to be imprecise and arbitrary.

For example, the Chinese equivalent of "individualism" has a negative connotation, meaning being selfish or self-centered. But Westerners value individualism positively. Here are some concordance lines from ACCE.

1. As many get richer, loyalty to the party and the authority (itself a product of propaganda) has been gradually replaced by the **individualism**, materialism and the pursuit of wealth.
2. Our traditional values of communalism and social harmony have been replaced by selfishness and **individualism.**
3. Always with foreign devils the same obsession with the self. Bourgeois **individualism.**
4. This world view has emerged not only with the celebration, almost fetishisation, of choice and **individualism**, but also from the collapse of vigorous working-class institutions that might have kept the wider culture more honest.
5. Money worship, hedonism and extreme **individualism** are spreading among some members of society. Looking for profit and forgetting righteousness, a lack of integrity, a loss of moral norms are occurring.

All the uses of *individualism* in the above concordance lines have negative connotations. *Individualism* in China English is described as materialistic in 1, selfish in 2, bourgeois in 3, fetishistic in 4, and hedonistic in 5. The dictionary could tell its users about the difference as shown below:

> **individualism** *noun* [U] **1** the belief that the freedom of individual people is more important than the needs of society or the government: *Capitalism encourages competition and individualism.* …
> **Note:** In China, *individualism* is considered to be completely contrary to the traditional value of collectivism. However, it is encouraged in Western culture.

In the above entry, a note is added to explain the connotational meaning of individualism in China. In addition, the illustrative example demonstrates clearly the core value of individualism in Western culture.

The headword *taipan* in OALD8 is defined as "a foreign person who is in charge of a business in China" without any labels or notes. However, it is an informal term used during the 19th and the beginning of the 20th century, and is becoming archaic. Therefore, a currency label *old-fashioned* should be added to guard the users against misusing it.

> **taipan** *noun* [C] (*old-fashioned*) a foreign person who is in charge of a business in China. …

Such a label would mean that the word is no longer used, but its counterparts in the real world exist, and we use a different name (*laoban*) for them. A similar label (*arch.*) can be added in the entry of *cheongsam* to indicate that the word form is still used, but some of its counterparts in the real world no longer exist.

Other subject field and register labels can be added in the entries for headwords of Chinese origin. For example, "*sports*" can be added in the entry for *t'ai chi ch'uan*, "*philosophy*" in the entry for *yang* and *yin*.

5.   Conclusion

To sum up, dictionaries of the future would become more regionalized with the globalization of the English language and the construction of corpora of world Englishes. The first drive is the theory of world Englishes. The existing EFL dictionaries for Chinese learners of English are Anglo-American centered no matter whether they are made in China or abroad. British or American English has long been regarded as the only 'standard' in China, while China English is seen as belonging on the periphery. The time has come, however, for long-standing beliefs and lexicographical practices to be questioned and changed. Dolezal (2006: 702) argues that "the lexicography of world Englishes creates a challenge to redefine, change, or reaffirm commonly held linguistic attitudes and ideas".

Under the theoretical framework of world Englishes, lexicographers should take varieties of English into account in their dictionaries. In the case of EFL dictionaries especially made for Chinese learners of English, it is advisable to include and describe words and use of China English in order to meet users' needs. It is generally acknowledged that English in China has distinctive linguistic and cultural features. The description of these in the dictionary can help its users adequately express ideas specific to Chinese culture.

Another drive is the availability of corpora of China English. Computer technology and corpora used in lexicography are generally regarded as an important "revolution in lexicographical resources and practices (Rundell 2010: 367)". Corpora of China English can provide lexicographers with fresh and revealing insights into the possible contents of an EFL dictionary. They enable lexicographers to analyze data on China English and make lexicographical decisions about lemma selection, defining styles and glosses and labels.

But the amount of China English to include in an EFL dictionary remains a matter for debate. A dictionary is always looked up to as a 'standard' or 'norm' by dictionary users, and they may find it difficult to accept an item from an English variety in a dictionary until they are provided with an adequate description of it by linguists and lexicographers. This therefore calls for linguists and lexicographers to conduct basic research into China English.

Fortunately we have seen concerted efforts of linguists, computer scientists and lexicographers in China. Thanks to the newly-developed corpora, linguists are able to investigate systematically the linguistic features of China English. Backed up by their research findings and corpus evidence, lexicographers are able to describe the linguistic and cultural features of words and phrases from China English adequately in their dictionaries.

5.   Acknowledgements

References

Allsopp, R. 1996. *Dictionary of Caribbean English Usage*. Oxford: Oxford University Press.

Benson, P. 1997. English Dictionaries in Asia: Asia in English Dictionaries. Bautista, M.L.S. (Eds.). 1997. *English is an Asian Language: The Philippine Context*: 125-140. Sydney: The Macquarie Library.

Bolton, K. 2006. *Chinese Englishes: A Sociolinguistic History*. Cambridge: Cambridge University Press.

Butler, S. 1997. Corpus of English in Southeast Asia: Implications for a Regional Dictionary. Bautista, M.L.S. (Eds.). 1997. *English is an Asian Language: The Philippine Context*: 103-124. Sydney: The Macquarie Library.

Cannon, G. 1988. Chinese Borrowings in English. *American Speech* 63: 3-33.

Delbridge, A. and S. Butler. 1999. The Macquarie Dictionary, its History and its Editorial Practices. *Lexikos* 9: 152-171.

Dolezal, F. 2006. World Englishes and Lexicography. Kachru, B.B., Y. Kachru and C.L. Nelson (Eds.). 2006. *The Handbook of World Englishes*: 694-708. Malden: Blackwell.

Du, R. and Y. Jiang. 2001. Jin ershinian "zhongguo yingyu" shuoping ("China English" in the past 20 Years). Waiyu jiaoxue yu yanjiu (Foreign Language Teaching and Researching) 33(1): 37-41.

Fang, C., F. Le and J. Cao. 2012. Zhongguo yingyu duibi yuliaoku de sheji, jianli he chutan (The Design, Establishment and Preliminary Study of a Comparable Corpus of China English). *Linguistic Research* 2: 113-127.

Gao, C. 2006. Shijie yingyu lilun yu zhongguo yingyu yanjiu zongshu (Literature Review on World Englishes Theory and China English). *Guowai waiyu jiaoxue* (*Foreign Language Teaching Abroad*) 4: 55-60.

Ge, C. 1980. Mantan you han yi ying wenti (Random Thoughts on Some Problems in Chinese-English Translation). *Fanyi Tongxun* (*Chinese Translator's Journal*) 2: 1-8.

He, D. and D.C.S. Li. 2009. Language Attitudes and Linguistic Features in the 'China English' Debate. *World Englishes* 28(1): 70-89.

Hu, X. 2007. Quanqiuhua huanjing xia "global English" de fazhan qushi ji dui zhongguo yingyu jiaoxue de qishi (The Development of "Global English" and its Implications for ELT in China in the Context of Globalization). *Guowai waiyu jiaoxue* (*Foreign Language Teaching Abroad*) 2: 16-23.

Jiang, Y. 2002. China English: Issues, Studies and Features. *Asian Englishes* 5(2): 4-23.

Kachru, B.B. 1985. Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. Quirk, R. and H.G. Widdowson (Eds.). 1985. *English in the World: Teaching and Learning the Language and Literatures*: 11-30. Cambridge: Cambridge University Press.

Kirkpatrick, A. 2002. *Englishes in Asia: Communication, Identity, Power and Education.* Melbourne: Language Australia.

Kirkpatrick, A. 2010. Researching English as a lingua franca in Asia: The Asian Corpus of English (ACE) Project. *Asian Englishes* 13(1): 4-18.

Kirkpatrick, A. 2011. English as an Asian lingua Franca and the Multilingual Model of ELT. *Language Teaching* 44(2): 212-224.

Li, D. and G. Liu. 2011. China Daily zhongguo yingyu yuliaoku gaikuang (An Overview of *China Daily* China English Corpus). *Yingyu guangchang* (*English Square*) 6(3): 58-59.

Li, W. 1993. Zhongguo yingyu yu zhongguoshi yingyu (China English and Chinglish). *Waiyu Jiaoxue Yu Yanjiu* (*Foreign Language Teaching and Researching*) 4: 18-24.

Li, W. 2007. Zhongguo yingyu xinwen baokan zhong de cicu (The Word Clusters in China English News Article Corpus). *Zhongguo Waiyu* (*Foreign Languages in China*) 4(3): 38-43.

Li, W. 2012. Yuliaoku biaoji yu biaozhu: yi zhongguo yingyu yuliaoku wei li (Corpus Markup and Annotation: China English Corpus as an Example). *Waiyu Jiaoxue Yu Yanjiu* (*Foreign Language Teaching and Researching*) 44(3): 336-345.

Liu, G. 2008. Xiandai daxing jidu yuliaoku de biaozhu fangfa: yi CEC zhengfu wenjian zi yuliaoku weili (The Annotation Methods of Modern Computer Corpora: A Case Study on the Annotation of the Government Document Corpus). *Shandong Waiyu Jiaoxue* (*Shandong Foreign Language Teaching Journal*) 126(5): 53-59.

Medgyes, P. 1992. Native or Non-native: Who's Worth More? *ELT Journal* 46(4): 340-349.

Mesthrie, R. 2010. *A Dictionary of South African Indian English.* Cape Town: University of Cape Town Press.

Pan, Z. 2005. *Linguistic and Cultural Identities in Chinese Varieties of English.* Beijing: Peking University Press.

Radtke, O.L. 2007. *Chinglish.* Layton: Gibbs Smith.

Rong, P. 1991. Chinese English is an Objective Reality. *Jiefangjun Waiyu Xueyuan Xuebao (Journal of PLA Foreign Languages Institute)* 1: 2-7.

Rundell, M. 2010. Taking Corpus Lexicography to the Next Level: Explicit Use of Corpus Data in Dictionaries for Language Learners. Zhang, Y. (Ed.). 2010. *Learner's Lexicography and Second Language Teaching: Proceedings of First International Symposium on Lexicography and L2 Teaching and Learning*: 367-386. Shanghai: Shanghai Foreign Language Education Press.

Sinclair, J. 1991. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sun, H. and J. Lv. 2010. Zhongguo yingyu yuliaoku yu yingyong (Corpus Establishment and Application of China English). *Journal of Henan Normal University* (*Social Science Edition*) 37(1): 219-221.

Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Wei, R. and J. Su. 2012. The Statistics of English in China. *English Today* 28(3): 10-14.

Wu, J. 2014. EFL shijiao xia gaoxiao yingyu jiaoshi de yuyan taidu yanjiu (A Study of the Language Attitude of College English Teachers from the Perspective of EFL). *Waiyu Jiaoxue* (*Foreign Language Education*) 35(2): 55-58.

Xia, L. and L. Zhai. 2016. Treatment of Entries with Chinese Characteristics in English Learner's Dictionaries: A Case Study of *Oxford Advanced Learner's Dictionary*. Eighth edition. *Kernerman Dictionary News* 24: 12-15.

Xu, Z. 2006. Rectifying "Chinese English". Hashim, A. and N. Hassan (Eds.). 2006. *Varieties of English in Southeast Asia and Beyond*: 283-291. Kuala Lumpur: University of Malaya Press.

Xu, Z. 2008. Analysis of Syntactic Features of Chinese English. *Asian Englishes: An International Journal of the Sociolinguistics of English in Asia/Pacific* 11(2): 4-31.

Xu, Z. 2010a. *Chinese English: Features and Implications*. Hong Kong: Open University of Hong Kong Press.

Xu, Z. 2010b. Chinese English: A Future Power? Kirkpatrick, A. (Eds.). 2010. *The Routledge Handbook of World Englishes*: 282-298. London: Routledge.

Yu, X. and Q. Wen. 2010. Zhongguo yingyu baozhang zhong pingjia xing xingrongci dapei de bentuhua tezheng (Collocation Patterns of Evaluative Adjectives of English in Chinese Newspapers). *Waiyu Jiaoxue Yu Yanjiu* (*Foreign Language Teaching and Research*) 254(5): 23-28.

Yule, H., A.C. Burnell and G.T. Kurian. 2006. *Cosmo Dictionary of Indian English*. New Delhi: Cosmo Publications.