# Researching Collocational Features: Towards China English as a Distinctive New Variety

## Jianli **Liang** and David C.S. **Li**

**Abstract:** This chapter shows that, in addition to isolated lexical items, collocational patterns represent an important type and source of innovation in China English (also known as 'Chinese English', 'CE' hereafter). Drawing on contrastive data analysis in two corpora – the Chinese English Corpus (CEC) and the British National Corpus (BNC) – we will demonstrate how evidence-based innovations are supported by statistically salient patterns of collocational features in CE. With the help of advances in information and communication technologies (ICT) which enable us to obtain a large amount of statistical information quickly, and to identify and prioritize any corpus-derived collocational patterns that we wish to scrutinize closely, such a process that used to be too cumbersome to handle manually is no longer difficult. Going beyond a descriptive account of the use of CE features that are embedded in collocations, this study argues that a comprehensive understanding and coverage of CE lexical features, and grammatical preferences to a lesser extent, cannot be obtained without attending to CE-specific collocational patterns, which constitute an important type and source of lexical and grammatical innovations. To the extent that the selection and codification of CE lexico-grammatical features are premised on systematic and reliable identification of collocational patterns, corpus-based methods hold the key to the future of research on CE towards its legitimation and recognition as a new variety of English.

**Keyword:** China English, World Englishes, collocational feature, corpus linguistics

## 1. Introduction

In the early 1980s, Braj Kachru (1982, 1992) made a cogent and ground-breaking argument that, following its global spread for over 200 years from the colonial to the postcolonial era, English should no longer be seen as homogeneous, let alone being seen as the exclusive property of the speakers of traditional English-L1 countries. Some forty years down the road, researchers from diverse first-language backgrounds in different parts of the world, working under the paradigm of World Englishes ('WE' hereafter), have taken big strides and greatly extended Kachru's original insights (e.g., Bolton, 2004; Bolton & Kachru 2006; Jenkins 2005; Kirkpatrick 2007, 2010; Kachru et al. 2006; McArthur 2002; Melchers & Shaw 2003; Schneider 2007; among many others). In an increasingly globalized world where 'native speakers' in traditional English-L1 countries are outnumbered by users of English as an additional language (EAL) or 'non-native speakers' of English, it is neither reasonable nor realistic to expect bi-/multilingual EAL speakers to adhere to the norms and standards of English that befit a 'native' variety, such as American or British English. To make local meanings in English without compromising their indigenous ethnocultural identities (Kirkpatrick 2007), EAL speakers engaged in ELF (English as a lingua franca) communication have no choice but to invoke or follow local lexical innovations by embedding them in their English output (Li 2011; Yang 2005). Such a need is especially pronounced when the topic of the conversation or text is of local relevance (Bamgbose 1998: 13; Jenkins 2006: 161). Lexical innovations reflecting and representing local meanings of socio-cultural import thus constitute significant evidence of and a compelling argument for a new variety of English, be it full-blown or in the making (Xu 2006: 283-285).

The phenomenal rise of China as a major player in the realms of world politics and the global economy after steadfastly adhering to the 'open door policy' since the early 1980s makes it self-evident that more and more Chinese people are connected to the rest of the world than ever, physically or via the Internet. All signs indicate that this momentum will be sustained in the near future. As an emerging variety of English with the largest number of speakers in the world, CE has attracted the attention and interest of many scholars who identify with the premises and goals of WE as a research paradigm, both in Mainland China and beyond (e.g., Adamson 2004; Bolton 2003; C. Gao 2008; Kirkpatrick 2007; Lo Bianco, Orton & Y. Gao 2009; Xu 2010; Yu, 2009). Many predicted that research on CE, being the variety with the largest number of speakers of English in the world, will likely have a significant role to play as it will have much to contribute to the theoretical underpinnings of and further development in WE research.

## 2. CE lexis: Pervious research

Research in CE to date has yielded valuable and significant insights on mainland Chinese tertiary students' awareness of CE as an emerging variety (e.g., Kirkpatrick & Xu, 2002) and made a strong case for CE as a new variety of English (see, e.g., Hu

2004; Lo Bianco, Orton & Y. Gao 2009; Xu 2010). Previous research typically points to isolated lexical words as evidence, focusing on words encoding and reflecting local culture and/or social practice. Two areas of investigation are discernible in the CE literature. The first investigates lexical borrowing of Chinese into English. For instance, Cannon (1987) classifies English words borrowed from Chinese into 19 semantic areas. *Soybean, tofu, oolong* and *wonton* are among many other examples cited. Similar studies are Cannon (1988), Yang (2009), Xiong (2007) and so on. The other kind of research goes one step further by examining CE lexical items whose connotation deviates from that in 'native' varieties. For example, Cheng (1992: 171) pointed out that the CE word *propaganda* has social values attached to it which make it sound neutral or even positive in the Chinese context, referring to 'news that is disseminated in a widespread way' in mainland Chinese newspapers and official documents. This observation is also attested in literary writing. For example, in *Death of a Red Heroine* (2000), a novel written by the CE writer Qiu Xiaolong, *propaganda* (p. 55) and *regime* (p. 361) were clearly used with a positive connotation, unlike their counterparts in 'native' varieties such as British or American English.

Two key issues arise when we evaluate the literature of studies on CE lexis to date: the depth and choice of research method. In view of the important implications of CE research in the domain of language education, plus a pressing need for exploring how CE research may inform the English language curriculum development in mainland China, many scholars point to the urgency of greater, concerted efforts to codify CE features systematically (e.g., Bolton 2003; Bolton & Graddol 2012; He & Li 2009; Kirkpatrick 2007; Xu 2010). Codification being an indispensable step toward gaining wider recognition, both internally and externally, its significance can hardly be underestimated, and yet progress to date has been slow and appears to be uncoordinated. One possible reason, among others, is that CE features to date either rely on examples based on the researchers' intuition, or are derived from data sets extracted from literary works in English written by Chinese authors. For example, Cannon (1988) and Yang (2009) identified CE lexis by comparing leading desk dictionaries, the results of which (i.e. whether a word is CE or not) are based on the lexicographers' intuition. There is no basis for checking to what extent the words listed are actually used by CE or other users. Xu's (2010) approach is clearly superior, in that CE lexical items were generated systematically from authentic data (newspapers and literary works in English, in addition to interview data with university students). One important highlight of Xu's research is that he analyses the language of Ha Jin, a novelist of Mainland Chinese origin who has written extensively in English. Partly because of the society Ha Jin grew up in and the education he received, he is a good example of an expert CE user. The evidence of the CE presented there is thus very much a focused case study. But as the body of data is relatively small, Xu (2010) suggests that more research needs to be carried out on more and wider data sources.

In sum, methodologically speaking, research on CE to date tends to be piecemeal and unsystematic, making it difficult to progress to the important and logical next step:

codification. After all, for a CE-specific lexical item or lexico-grammatical feature to be reliably incorporated into a CE-informed national ELT curriculum, we need solid evidence that such an item or feature is actually used by a significant percentage of members of the CE community at large. As the basis for selection, intuition-based methods are practically unreliable and potentially objectionable. How do we gain access to real-life CE texts used for authentic communication purposes? In the past, processing massive amounts of language data manually (e.g., exhaustive classification and cataloguing through cut and paste) was simply not a feasible option. From the 1990s onwards, that problem has largely been obviated by significant breakthroughs in computer science research as well as giant leaps in ICT technologies in the last two decades. Indeed, once a keyword search is defined, linguistic patterns in terms of collocations and frequencies can be generated instantly with the help of powerful software. Such patterns are typically just a few clicks away. Regarding the question 'How is English actually used by educated CE users?', technologically and methodologically we have the know-how today to fill the knowledge gap involving the identification of CE-specific lexis and lexico-grammatical features, one that can be filled by large-scale, data-driven and corpus-based research techniques.

Our goal in this chapter is therefore twofold in essence: (a) to demonstrate why identifying CE-specific collocational patterns and lexical features is crucial towards legitimizing CE as a new variety of English; and (b) to make a case for corpus-based research methods as the preferred tool or platform given their tremendous potential in discovering widely shared collocational patterns and lexico-grammatical features among CE users.

## 3. CE collocations

### 3.1 Collocation

The idea that we know the meaning of a word best, not by examining it in isolation, but by observing other words with which it co-occurs is usually ascribed to Firth in his well-known maxim "You shall know a word by the company it keeps" (Firth 1968, p. 179), which emphasizes the importance of collocation studies in the description of language use. Discovering the extent of word associations and how words are used in terms of their collocational patterns has been an important area of study in corpus linguistics since the 1960s. With the growth of corpus linguistics, there has been increasing interest in collocation studies. Today, nobody seriously interested in the meaning and use of language can ignore tendencies of word co-selection which are evident in linguistic patterns, the investigation of which has always been in the core interest of the field of applied linguistics.

It should be made clear, however, that some other scholars use different terms to refer to more or less the same collocational patterns. Some examples of these are 'lexicalised sentence stems' (Pawley & Syder 1983), 'chunk' (Sinclair 1991), 'lexical phrase'

(Nattinger & De Carrico 1992), 'lexical bundle' (Biber et al. 1999), 'cluster' (Scott 2007), 'formulaic sequence/expression' (Wray 2000), 'phraseology' (Renouf 1992; Stubbs 2002), 'chunks' (Schmitt 2010), 'multi-word units' (Greaves & Warren, 2010), among others. In this chapter, all these terms are treated as synonymous, although the word 'collocation' will be used throughout, and it is defined as:

> a group of words which occur repeatedly in a language. These patterns of co-occurrence can be grammatical in that they result primarily from syntactic dependencies or they can be lexical in that, although syntactic relationships are involved, the patterns result from the fact that in a given linguistic environment certain lexical items will co-occur. (Carter 2012: 62)

### 3.2 CE collocation: Blood Moon or Red Moon?

Collocation, as reflected in the co-occurrence of word choice, is an important part of vocabulary research. For instance, Halliday (1966: 149) observes that, whereas *powerful* and *dark* may be construed as synonyms of *strong* and *black*, respectively, the former normally do not collocate with *tea* or *coffee* (viz.: *strong tea*, *\*powerful tea*; *black coffee*, *\*dark coffee*), even though semantically the less idiomatic-sounding coupling has very similar meaning as that of the preferred coupling. Examples featuring the co-occurrence of words like *strong* and *tea*, *black* and *coffee* are generally referred to as 'collocations' (Halliday 1966: 149). This classic example shows that many apparent synonyms, such as *strong* and *powerful*, may have characteristically different collocations, and so the co-occurrence relationship of a given word with other words (i.e., collocations) is an important part of its meaning and usage. Replacing the usual collocation with a synonym, (e.g., substituting the adjective *powerful* for *strong [tea]*, or *dark* for *black [coffee]*) is grammatically possible but hardly ever done. *Powerful tea* and *dark coffee* are therefore unusual collocations (compare: *dark-roasted coffee*, and '*The Powerful Tea for Athletes and All!*', the latter being an advert for a tea product, https://www.kickstarter.com/projects/matebros/matebros-the-powerful-tea-for-athletes-and-all).

In earlier studies of the use of English in China, the unusual collocation of a given word (e.g., *new innovation*, *family relatives*, Pinkham 2000: 26) not found or shared in standard, 'native' varieties of English was used as justification to label that collocation as non-standard, an error in need of correction. This raises the question, whether users of a new variety of English can evolve their own preferred collocations. Consider the following examples:

1) People are expected to enjoy the *Red Moon* or the total lunar eclipse on October 8, 2014 in Shanghai if the weather is fine. (Source: *Shanghai Daily*, 6-10-2014)
2) The last time the city saw the *Red Moon* was 3 years ago. (Source: *China Daily*, 7-10-2014)
3) A total lunar eclipse has been visible across much of the Americas and Asia,

resulting in a dramatic "*Blood Moon*". (Source: *BBC website*, 10-10-2014)
4) *Blood Moon* appears after lunar eclipse. (Source: BBC website, 10-10-2014)

Examples 1) to 4) were extracted from news reports in China and UK. Both reports are concerned with the same story, an unusual astronomical phenomenon: a special reddish color of the moon triggered by a total lunar eclipse. Interestingly, whereas in China the word *red* was used to pre-modify *moon*, in British English in reference to its color the word and imagery *blood* was used instead. In linguistic studies, examples such as these constitute sound evidence of locally preferred collocations. In contradistinction to *blood moon* in BrE, *red moon* may be regarded as the preferred collocation in CE. In accordance with a widely accepted premise in WE research, that lexical variants (e.g., *hutong* vs. *street*) should be viewed as difference rather than deficit, there is room for arguing that the collocation *red moon* as found in the Chinese context reflects CE users' collocational preference rather than being non-standard merely because it deviates from the preferred collocation in an inner circle variety. [The discussion on Red Moon vs. Blood Moon is an interesting one, but it seems that it's only based on four examples, instead of their respective corpora. More examples, or statistics could be expected. As a general reader, I would also expect that there are cases of 'blood moon' used in the Chinese English corpus.]

Recognition of lexical variants may also be found in creative works. For example, in a popular film produced by the American company DreamWorks Studios in 2008, *Kung Fu Panda*, in the original sound track (English) the term of address *master* and its translation equivalent in Mandarin *shifu* (Pinyin: *shīfu*) are both used. Their meaning is basically identical; both are used by learners (a panda, a snake, a crane, a tortoise, etc.) of Kung Fu to address their teacher. The actual, interchangeable use of these two words alternately, in free variation by all the characters in the film targeting audiences worldwide, suggests that lexical variants and alternations involving Chinese-specific words today are quite acceptable.

In sum, collocational variants of CE should be investigated while researching lexical variants. If *master* and *shifu* are used interchangeably, there is no reason why *red moon* cannot be recognized and accepted as a legitimate 'collocational variant' (see Liang 2015: 306; cf. "phraseological variation", Greaves & Warren 2010: 218) of *blood moon*, even though the latter is preferred in an inner circle variety (BrE, also attested in AmE such as *New York Times*).

### 3.3 Corpus-based research into CE collocation

To identify overt collocational features like 'red moon' in CE and 'blood moon' in BrE and AmE comprehensively and systematically, the manual search-and-tag method is clearly inefficient, unproductive and error-prone, partly because observations of linguistic patterns are generally accidental and unrepeatable. Fortunately, advances in digital computation technologies in the last two decades have made it possible to

conduct large-scale searches for target words or chunks in multi-million-word corpora and databases in a fraction of a second (Liang 2015: 281). As Sinclair (1991: 100) pointed out, "the language looks rather different when you look at a lot of it at once"; we should therefore employ corpus-based tools and methods to identify CE collocational features. In other words, corpus-based research methods hold the key to the systematic discovery of covert and hitherto unknown CE collocations. At the operational level, there are two important characteristics that are not found in manual search-based studies and observations, 'co-occurrence' and 'high frequency', which may be defined as follows:

**Co-occurrence:** a particular mode or pattern in which words occur together. Embedded meaning is created by virtue of multiple occurrences of these patterns, representing an important part of the meanings of these words. The semantic profile of one word is established by the co-occurrence with other words. Within corpus linguistics, for a co-occurrence relationship to be called a collocation, the number of times of recurrence varies from twice (Sinclair 1991) to three times (Clear 1993: 277).

**High frequency:** the number of co-occurrences must be sufficiently high to qualify them as 'fixed' expressions. The cut-off point for 'high' or 'low' frequency varies, depending not only on the size of the corpus and the purpose of research, but also on individual researchers' personal preference to some extent.

Any analysis of typical choices of co-occurrence of words depends on frequency analysis. The very mention of a choice being 'typical' or 'rare' implies that, under given circumstances, it happens more or less often than other choices. Obviously then, in order to identify CE collocational patterns, we need to compare whether those patterns occur in at least one reference corpus. 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.

*3.3.1 Data and Method*

Using the 37-million China English Corpus (CEC) (37 million words, W. Z. Li 2010) as the target data with the British National Corpus (BNC_web, written part: 90 million words) serving as the reference corpus, we made a comparison of high-frequency collocations. Below we will illustrate a number of unique collocations which appear to be only used by CE but not BrE users, which we will characterize as 'innovative collocations'. By using the word 'innovative', the peculiarity of these collocations is emphasized. One very important premise is that, the absence of any tokens in one or more inner circle varieties does not constitute evidence of the type (word or phrase) in question being unacceptable or problematic, let alone erroneous.

*3.3.2 Result: the case of* **all-round**

Frequency information of the commonly encountered word *all-round* is easily obtained, and its collocations may be checked instantly where necessary, using the freeware corpus analysis toolkit AntConc 3.2.4 (Anthony 2011). Table 1 shows the number of tokens of *all-round* in CEC and BNC, including the average number of tokens per million words:

Table 1: Frequency information of the word *all-round*

|  | **CEC** | **BNC** |
| --- | --- | --- |
| Occurrence (tokens) | 543 | 251 |
| Tokens/per million words | 14.68 | 2.86 |

As the ratio shows, the word *all-round* is used more often (approximately 5 times) in CE, suggesting a stronger inclination among CE users to use it as a pre-modifying adjective. Extending the search to the R1 collocations of *all-round*, it was found that three collocates constitute 45% (244/543) of all the R1 patterns of *all-round* in CEC (Table 2).

Table 2: Top 3 collocations of *all-round* in CEC

|  | **Occurrence (actual tokens)** |
| --- | --- |
| all-round way(s) | 183 |
| all-round development(s) | 49 |
| all-round cooperation(s) | 12 |
|  | Sum = 244 |

(*span = R1: the AntConc search sets the span as the first word to the right)

These three high-frequency usage patterns are clearly distinct. Of particular interest is the first collocation, *all-round way(s)*, which is found across all the genres in the CEC, suggesting that it is widely accepted in different fields and everyday lives rather than being specific to a restricted number of topics. Of the 183 tokens of the R1, two-word collocation *all-round way(s)*, 178 occur in the longer pattern *in an all-round way(s)*. A subsequent search in the BNC corpus did not yield any occurrence of this collocation. On the basis of this finding, we propose that the collocational pattern *in an all-round way(s)* is one example of a distinct CE collocation. Below are five CEC text segments in which this collocational pattern is embedded:

5) In the meantime, in order to promote regional sustainable development, construction of some 100 ecological demonstration areas, mainly at the county level, has been started *in an all-round way.* (domain: applied social science)

6) Science, technology, education and other social undertakings developed *in an all-round way.* (domain: world affairs)

7) The quality of the teaching staff should be improved *in an all-round way.* (domain: social science)

8) We will fully implement the Party's education policy by strengthening moral education and promoting competence-oriented education to ensure that students develop *in an all-round way.* (domain: education)

9) The only choice is to accelerate opening, using international resources *in an all-round way* and exploring the international market. (domain: economics)

*All-round,* unlike some other words specific to the social realities in China (e.g., *hukou system,* which is usually translated as 'household registration system'), by no means denotes a Chinese-specific reality. Thus, the sheer number of tokens of *in an all-round way* – 178 tokens in CEC but none in BNC – is strongly suggestive of CE users' collective preference for this collocational pattern, as evidenced by its relatively high frequency.

It should be noted that the hyphenated compound adjective *all-round* is used in BrE too, but its collocations are rather different. Following the same procedure for investigating frequency, we further compared the collocational patterns involving *all-round*. The top 8 collocations in CEC and BNC are listed in Table 3.

Table 3: Collocations of *all-round* in BNC and CEC (in decreasing order of frequency)

| BNC | | CEC | |
|---|---|---|---|
| 1. | all-round SPORTSMAN | 1. | all-round WAY |
| 2. | all-round PERFORMANCE | 2. | all-round DEVELOPMENT |
| 3. | all-round VISION | 3. | all-round COOPERATION |
| 4. | all-round PLAYER | 4. | all-round PROGRESS |
| 5. | all-round EXCELLENCE | 5. | all-round MANNER |
| 6. | all-round TEAM | 6. | all-round IMPROVEMENT |
| 7. | all-round STRENGTH | 7. | all-round PARTNERSHIP |
| 8. | all-round DEVELOPMENT | 8. | all-round CONSIDERATION |

*(span = R1)

Six of the eight CEC collocations (no. 1, 3, 4, 5, 6, 7, 8, as highlighted) are unique in CE and not found in the entire BNC. One example of each of these distinct collocations is presented below.

10) Then he expressed his wishes to launch *all-round cooperation* with Tongji University. (domain: education)
11) We need to improve the people's lives, increase social harmony and promote *all-round progress* in socialist economic, political, cultural, social and ecological development. (domain: politics)
12) Let us work diligently and carry forward in an *all-round manner* the cause of reform and opening up and the socialist modernization drive into the new century. (domain: politics)
13) China and Belgium decided on Monday to upgrade their relations to an *all-round partnership* of friendship and cooperation. (domain: world affairs)
14) The environmental effects shall be evaluated in an objective, open and impartial manner, with an *all-round consideration* given to the possible effects on the various environmental factors and on the ecological system, which is composed of the factors, after a plan is implemented or a project is constructed,

providing a scientific basis for decision-making. (domain: environmental protection)

To the extent that the examples 10) to 14) are grammatically and semantically well formed, it would be groundless to label these high-frequency phrases as problematic or dismiss them as 'bad English' or errors, just because they are not used by speakers/writers of 'native speakers' of inner circle varieties. Quite the contrary, collocations such as *in an all-round way(s)* are perfectly well-formed and should be accepted as part of CE collocations. Our main argument here is that when a word or collocational pattern is uncommon among speakers/writers of one or more inner circle varieties of English, it would be premature to dismiss it as an error solely on account of its non-existence or apparent deviation from an inner circle variety. Rather, such cases should be examined in their own right, and corpus-based information such as its frequency and range among users of a variety provides compelling evidence of its well-formedness in that variety.

Two collocations are used by both BNC and CEC in Table 3, i.e., *all-round development(s)* and *all-round improvement(s)*, which may be regarded as evidence of their acceptability among BrE users. On closer scrutiny, however, we found that the same collocation is used rather differently in the CEC and BNC. Three examples of the collocation *all-round development(s)* in each corpus are presented below for comparison and contrast:

Text segments containing *all-round development(s)* in CEC (3 out of 49):

15) The government has carried out the strategy of seeking *all-round development* to turn the city into an important iron and steel, energy and vanadium-titanium industrial center, as well as a high-quality sub-tropical agricultural and sideline production base. (domain: society)
16) The two sides will exchange views on bilateral relations as well as on issues of mutual interest in the regional and global arena with the aim of promoting comprehensive, *all-round development* of relations. (domain: political issue)
17) We should continue to increase investment in agriculture, promote *all-round development* of agriculture and the rural economy and work to increase farmers' incomes. (domain: agriculture)

Text segments containing *all-round development(s)* in BNC (3 out of 6):

18) The work of Norman Shneidman has been of some value in highlighting the way in which sport in Soviet society is 'directed towards the *all-round development* of the human individual'. (domain: leisure)
19) However, the term general can be conceived of in a third way, in terms of the *all-round development* of the individual. (domain: social science)
20) 'We are looking at the *all-round development* of the child, not just narrow

academic achievement,' added Ms Robinson. (domain: social science)

Notice that although the same collocation is used, it is used with regard to different entities. Whereas in BNC, *all-round development(s)* is used in reference to people, in CEC, *all-round development(s)* is typically used to pre-modify things or abstract entities (e.g., economy, agriculture, society). In terms of its semantic scope and range, therefore, there is some evidence that whereas BrE users tend to use *all-round development(s)* to refer to people, CEC users tend to use it to refer to abstract entities.

In addition to speedy access to statistical information such as frequency and collocation, another methodological edge of corpus-based research is the quick revelation of any embedded linguistic feature(s) of a given collocation, as illustrated in the *Key Word in Context* (KWIC) window (Figure 1).

Figure 1: KWIC of *in an all-round way* in CEC



As shown in Figure 1, some of the KWIC-induced insights include: (a) the domains or specific topics in which the target node word occurs (e.g., sports & health, politics, economics, leisure), (b) the lexemes and their derivatives which frequently co-occur to the left and right of the node word (e.g., *develop(ed/ing/ment)*), and (c) possible linguistic 'partners' with which it contracts a collocational relationship, which are easily displayed alphabetically from a to z (realized by setting the "*span*" and "*sort*" function in the AntConc toolkit).

## 4. Discussion

We hope to have made a convincing case why *in an all-round way(s)* should be

recognized as a legitimate CE collocation, and that there should be no more doubt about whether it should be edited out or modified to conform to the usage pattern of an inner circle variety like BrE. Research on CE lexical features should focus not only on the exotic or uncommon words, but also on everyday vocabulary and the collocational patterns that go beyond the confines of a single word. CE-specific words such as *hutong, hukou, wonton* are certainly important. This should continue as and when they are detected, but this should be complemented by research on usage patterns involving ordinary, high-frequency words as well. As shown in CE users' use of *all-round* above, this compound adjective is by no means specific to CE vocabulary (e.g., it also exists in BrE). How is it actually used by CE as opposed to BrE users, and in what collocations is it embedded in CE contexts? Are there lexical features and collocational patterns that are specific to CE, regardless of whether the words in question are common or uncommon? Without corpus-based research techniques, it may not be obvious how these questions could be satisfactorily addressed. This is why, to answer these questions, collocational patterns must be included in further research on CE.

There are at least two advantages for conducting corpus-based research into CE collocational features, namely retrievability and quantification. Retrievability refers to the possibility of getting the same results from corpus data, together with the repeatable procedures that would minimize the undue interference of personal factors (e.g., the researchers being anxious or exhausted). Being electronically retrievable is one natural advantage of corpus linguistics. Texts stored in designed format (raw or annotated) are ready-made for searching and sorting. This characteristic allows researchers to nominate search items or generate hypotheses without having any specific search items in mind. For example, the search item *all-round* was originally not self-nominated by us. As a collocation, its distinctive nature in CE emerged only after its statistical salience (i.e., high frequency) was confirmed through repeated searches, especially in comparison with British English. It was after further investigation that the CE patterns of *all-round*, i.e., innovative collocations such as *in an all-round way(s), all-round progress* were identified. Compared with intuition-based evidence that tends to be subjective and impressionistic, corpus-based procedures are more promising in that they yield results that are far more objective and robust.

Objective (or inter-subjective) corpus-based research methods are not mutually exclusive with subjective observations grounded in intuition. The two are complementary and must be so in order that as broad a range of research questions as possible can be addressed by linguistics. It is only with the help of large-scale, well-conceived corpus-based research that otherwise obscure or non-transparent linguistic patterns such as high-frequency collocations have a good chance of seeing the light. Advances in ICT have obviated the cumbersome task of manual tallying, and replaced it with user-friendly and efficient research techniques. Corpus-based research has the unrivaled capacity to process huge amounts of language data, to re-run any data-processing procedures at will, and to yield more objective results. With good potential to combine the advantages of quantitative and qualitative data analysis, corpus-based

research methods may well be a promising and productive way to explore the nuts and bolts of CE, including, for example, variation among CE users depending on their L1 ('dialect') backgrounds, professional orientations, academic disciplines, and the like. All in all, towards winning greater recognition of CE-specific lexico-grammatical features and patterns, we believe corpus-driven research will prove to be an indispensable tool.

## 5. Conclusion

Since the new millennium, the enhancement of China's national economic strengths worldwide has gradually improved its soft power globally, allowing it a greater say in the international community. English is now no longer a symbol of privilege for its users; rather, it has become a tool for identity-creation in spoken discourse as well as various written, e-communication platforms (e.g., using CE as ELF when interacting with people from other language backgrounds), fostering the hope of achieving recognition and respect for the nation. In this global context, comprehensive research into the distinctive features of CE has, and will continue to contribute to, a better understanding of CE users as they are engaged in communication in English, both intra- and internationally. With smart phones and other electronic gadgets becoming more and more popular and accessible, people involved in such communication need to recognize clearly that the cultural and identity-related factors in international communication are as important as language itself, if not more so. CE collocational features make up one such factor that deserves to be recognized. An important aim of this chapter is to argue for including collocational patterns in the analysis of lexical and pragma-linguistic features in CE, with a view to extending the depth of such analyses, refining the description of such features, and providing a clearer research plan toward the systematic discovery and codification of lexico-grammatical features in CE. A corpus-based approach towards better understanding what China English consists of should be promoted, and the lingua-cultural character of CE as a semiotic, meaning-making system at the disposal of its users can be gained by a detailed and empirically-grounded analysis of its distinctive linguistic features. From discovering statistically salient collocational patterns to their eventual inclusion in the nation's English language teaching (ELT) curriculum, it is our belief that corpus-based methodologies are destined to play a crucial, instrumental role in advancing the national research agenda and further development of CE as a member of the family of World Englishes.

General comments: A very interesting chapter, contributing to the literature of corpus based studies in Chinese English. The 'all-round' example is well-chosen and clearly elaborated.

**References:**
Adamson, B. (2004). *China's English: A history of English in Chinese education.* Hong Kong: Hong Kong University Press.
Anthony, L. (2011). AntConc (Version 3.2.4) [Computer Software]. Tokyo: Waseda

University.

Bamgbose, A. (1998). Torn between the norms: Innovations in world Englishes. *World Englishes*, 17(1), 1-14.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

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

Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (eds.), *The Handbook of Applied Linguistics* (pp. 367–396). Oxford: Blackwell Publishing.

Bolton, K. & Kachru, B. (2006). *World Englishes: Critical Concepts in Linguistics*. London: Routledge.

Bolton, K. & Graddol, D. (2012). English in China today. *English Today*, 28(1), 3-9.

Cannon, G. (1987). Dimensions of Chinese borrowings in English. *Journal of English Linguistics,* 20(2), 200-206.

Cannon, G. (1988). Chinese borrowings in English. *American Speech*, 3-33.

Carter, R. (2012). *Vocabulary: Applied linguistic perspectives*. London: Routledge.

Cheng, C. (1992). Chinese varieties of English. In B. B. Kachru (ed.), *The other tongue: English across cultures (2ⁿᵈ ed.)*, Urbana and Chicago: University of Illinois Press.162-177.

Clear, J. (1993). From Firth principles: Computational tools for the study of collocation. In M. Baker, G. Francis & E. Tognini-Bonelli (eds.), *Text and technology: in Honour of John Sinclair* (pp. 271-292). Amsterdam: John Benjamins.

Firth, J. R. (1968). A synopsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected papers of J.R. Firth 1952-1959* (pp. 168-205). Harlow: Longman.

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

Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units. In A. O'Keeffe & M. McCarthy (Eds.), *Routledge handbook of corpus linguistics* (pp.212-226). Routledge.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.), *In memory of J. R. Firth.* (pp. 148-162). London: Longman.

He, D. & Li, D. C. S. (2009). Language attitudes and linguistic features in the 'China English' debate. *World Englishes*, 28(1), 70-89.

Hu, X. (2004). Why China English should stand alongside British, American, and the other 'world Englishes'. *English Today*, 20(2), 26-33.

Jenkins, J. (2003). *World Englishes*. London: Routledge.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly,* 40(1), 157-181.

Kachru, B. B. (1982/1992). *The other tongue: English across cultures.* Chicago: University of Illinois Press.

Kachru, B. B., Kachru, Y. & Nelson, C. L. (2006). *The Handbook of World Englishes.* Oxford: Blackwell Publishing.

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

Kirkpatrick, A. (2010). *The Routledge handbook of world Englishes*. Abingdon: Routledge.

Kirkpatrick, A. & Xu, Z. (2002). Chinese pragmatic norms and 'China English'. *World Englishes,* 21(2), 269-279.

Kirkpatrick, A., & Xu, Z. (2012). *Chinese rhetoric and writing: An introduction for language teachers.* The WAC Clearinghouse and Parlor Press.

Leech, G. N. (2002). *The importance of reference corpora.* Keynote speech at Hizkuntza-corpusak. Oraina eta geroa. (2002-10-24/25). Retrieved from: http://www.corpus4u.org/forum/upload/forum/2005060301260076.pdf. Accessed 10 July 2014.

Li, D. C. S. (2011). Lexical gap, semantic incongruence, and medium-of-instruction-induced code-switching: Evidence from Hong Kong and Taiwan. In Eric A. Anchimbe & Stephen A. Mforteh (eds.), *Postcolonial linguistic voices: Identity choices and representations* (pp. 215-240). Berlin: Mouton de Gruyter.

Li, W. (2010). China English Corpus. Project funded by National Planning Office of Philosophy and Social Science of People's Republic of China.

Liang, J. L. (2015). *Collocational features of China English: A corpus-based contrastive study.* PhD dissertation, The Hong Kong Institute of Education.

Lo Bianco, J., Orton, J. & Gao, Y. H. (2009). *China and English: Globalisation and the dilemmas of identity.* Bristol: Multilingual Matters.

McArthur, T. (2002). *The Oxford guide to World Englishes*. Oxford: Oxford University Press.

Melchers, G. & Shaw, P. (2003). *World Englishes.* London: Arnold.

Nattinger, J. R., & De Carrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Qiu, X. (2000). *Death of a Red Heroine*. New York: Soho Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.

Pinkham, J. (2000). *Zhongshi yingyu zhi jian [Lessons Drawn from Others' Mistakes of Chingish]*. Beijing: Foreign Language Teaching and Research Press.

Renouf, A. (1992). What do you think of that? A pilot study of the phraseology of the core words of English. In G. Leitner (Ed.), *New directions in English language corpora: methodology, results, software development* (pp. 301-317). Berlin: Mouton de Gruyter.

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual.* Basingstoke: Palgrave Macmillan.

Scott, M. (2007). *Homing in on the text-initial cluster*. Paper presented at the Aston corpus symposium. Retrieved from http://acorn.aston.ac.uk/symposium. html. Accessed 12 September 2013.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Stubbs, M. (2002). *Words and phrases: corpus studies of lexical semantics.* Oxford:

Blackwell Publishers.

Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied linguistics*, 21(4), 463-489.

Xiong, X. (2007). Chinese English. In M, Nakano, (ed.). *World Englishes and Miscommunications.* (pp. 141-166). Tokyo: Waseda University International.

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

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

Yang, J. (2005). Lexical innovations in China English. *World Englishes*, 24(4), 425-436.

Yang, J. (2009). Chinese borrowings in English. *World Englishes*, 28(1), 90-106.

Yu, X. (2009). *Nativized lexical use in China's English newspapers*. PhD dissertation, Najing University, China.