

COLLOCATIONAL FEATURES OF
CHINA ENGLISH:
A CORPUS-BASED CONTRASTIVE STUDY

中國英語搭配特點：
基於語料庫的對比研究

by

LIANG Jianli

A Thesis Submitted to
The Hong Kong Institute of Education
in Partial Fulfillment of the Requirement for
the Degree of Doctor of Philosophy

April 2015



The Hong Kong
Institute of Education Library

For private study or research only.
Not for publication or further reproduction.

UMI Number: 3707390

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3707390

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346



The Hong Kong
Institute of Education Library

For private study or research only.
Not for publication or further reproduction.

STATEMENT OF ORIGINALITY

I, LIANG Jianli, hereby declare that I am the sole author of the thesis and the material presented in this thesis is my original work except those indicated in the acknowledgement. I further declare that I have followed the Institute's policies and regulations on Academic Honesty, Copyright and Plagiarism in writing the Thesis and no material in this thesis has been published or submitted for a degree in this or other universities.



Thesis Examination Panel Approval

Members of the Thesis Examination Panel approve the thesis of LIANG Jianli
defended on (October/14/2014).

Supervisors

LI Chor Shing, David
Professor
Department of Linguistics and
Modern Languages Studies
The Hong Kong Institute of Education

WANG Lixun
Associate Professor
Department of Linguistics and
Modern Languages Studies
The Hong Kong Institute of Education

MA Qing
Assistant Professor
Department of Linguistics and
Modern Languages Studies
The Hong Kong Institute of Education

Examiners

HABERLAND Hartmut
Professor
German Language and
Sociolinguistics of Globalisation
Roskilde University, Denmark

MOODY Andrew
Associate Professor
The Department of English
University of Macau
Macau SAR

CHIN Chi On, Andy
Assistant Professor
Department of Linguistics and Modern
Languages Studies
The Hong Kong Institute of Education

Approved on behalf of the Thesis Examination Panel:

Chair, Thesis Examination Panel
LO Sing Kai
Professor
The Dean of Graduate School
The Hong Kong Institute of Education



ABSTRACT

COLLOCATIONAL FEATURES OF CHINA ENGLISH: A CORPUS-BASED CONTRASTIVE STUDY

LIANG Jianli

The Hong Kong Institute of Education

Abstract

Research has shown that the form of English used by people in Mainland China (referred to as *China English* in this thesis; hereafter CE) is characterised by many distinctive features, both explicit and implicit, some of which are categorised as grammatical errors or non-standard English, while others as lexical items distinct to Chinese cultures, both of which are, of course, to be expected.

However, deviations that emerge in the form of co-occurrence of words or word categories are independent of these two groups. They often involve features at a structural level, i.e. collocation, colligation, semantic preference and semantic



prosody (Sinclair, 2004, p. 39), which are broadly referred to as ‘collocational features’ in this thesis. Such collocational features play a central role in the description of CE features, since they are integral to the language (rather than errors which are in need of correction), but also because CE users’ L1 identity and ownership of English should both be respected.

The goal of this study is to present the findings of a systematic empirical investigation of CE features in contra-distinction to other varieties of English in terms of collocations, words that co-occur and co-select in a distinctively Chinese way. The corpus used as a reference for contrast is British National Corpus, which represents British English. Departing from a World Englishes (hereafter WE) perspective, this study is a corpus-based contrastive analysis of some selected collocational features of CE and standard varieties of English in the inner circle. CE collocational features that are statistically salient and yet frequently dismissed as resilient acquisitional problems may pose a challenge to mutual understanding in a ‘world village’ where English is increasingly used as a *lingua franca*. Furthermore, since CE is predicted to play a more important role in the foreseeable future (Lo Bianco, Orton & Gao, 2009; Xu, 2010), the cultural identity embedded in it should be respected and accepted (Y. Gao, 2004; D. C. S. Li, 2010). In view of such predictions, this thesis aims to provide an adequate, data-driven description of the CE phenomena, followed by



conceptualisation and explanation through contrastive studies.

The following research questions will be addressed:

1. Are there distinctive collocational features in the English used by Chinese?
2. Are there colligational preferences in CE?
3. Are there semantic and prosodic preferences in CE?
4. Is it feasible to formulate a theoretically sound framework for analysing collocational features of China English based on the answers to (1)-(3)?

Taking advantage of statistical information to help identify and prioritise the corpus-derived collocational items, a process which is traditionally too cumbersome to handle by manual examination, this study analyses and provides a descriptive account of the use of CE in terms of collocation (Sinclair, 2004). The results reveal that while CE-specific collocational patterns constitute an important type of collocational innovations in CE, they also allow us to look into the important process whereby CE is evolving into a new variety of English. Collocational features constitute a significant type of evidence and argument when assessing to what extent a new variety of English is emerging.



ACKNOWLEDGEMENTS

First and foremost, I wish to express my deepest thanks to my supervisors, Prof. LI Chor Shing, David, for his unreserved professional guidance, careful reading and valuable comments. His inspiration, patience and heartfelt encouragement are very important to me. He is a conscientious and responsible scholar who is highly qualified in guiding me in my study and thesis writing. Many of the ideas presented here also benefited greatly from discussions on an early version with Prof. Andy KIRKPATRICK, whose suggestions have been a good spur and resource. My gratitude also goes to my co-supervisors WANG Lixun, MA Qing, and XU Zhichang, for their critical comments and invaluable suggestions. I alone am responsible for any remaining inadequacies of the thesis.

I would also like to express my gratitude to my friends Dr. WANG Danping, Ms. QIAN Zhen for their kind advice and timely help during my study.

I wish to give my deepest appreciation to my beloved parents who have continually provided me with everlasting love, unconditional encouragement for decades. Last but not least, I own a heart-felt debt to my husband Mr. ZHANG Zhiwen and my daughter ZHANG Kexin. Without their support, my pursuit of a PhD degree would not have been so enjoyable. Their love, encouragement, and tolerance have made my life richer and more meaningful, which embody the true spirit of bliss to me.



TABLE OF CONTENTS

Statement of Originality	ii
Thesis Examination Panel Approval	iii
Abstract	iv
Acknowledgements	vii
List of Abbreviations	xii
List of Tables	xiii
List of Figures	xvi
 Chapter 1: Introduction	 1
1.1 Globalisation and the global spread of English	1
1.2 Organization of the thesis	9
 Chapter 2: Literature review	 11
2.1 The World Englishes perspective	12
2.1.1 Models of Englishes	12
2.1.1.1 ENL/ESL/EFL distinctions	13
2.1.1.2 Inner/outer/expanding circles	17
2.1.1.3 Global language system: peripheral/central/ supercentral/hypercentral languages	21
2.1.1.4 Summary	24
2.1.2 Norms and standards	24
2.1.2.1 Standard Englishes	24
2.1.2.2 Non-standard Englishes	36
2.1.2.3 Summary	39
2.1.3 Standards and CE	42



2.1.3.1	Awareness of linguistic variability	43
2.1.3.2	Linguistic prejudice	45
2.2	Definition of China English	46
2.3	The China English debates	48
2.3.1	Debates in Mainland China	50
2.3.2	Debates beyond Mainland China	56
2.4	Learning and teaching English in China	59
2.4.1	Thirty years of the CET4 test	59
2.4.2	The non-native teachers	69
2.4.3	CE versus interlanguage	74
2.4.4	The status of English in China	87
2.4.5	Research on attitudes towards CE	91
2.5	Corpus linguistics and its use in research on English varieties	93
2.5.1	Real language and the corpus revolution	93
2.5.2	Frequency: key feature	98
2.5.3	Types of corpus	99
2.5.4	Approaches in corpus linguistics research	101
2.6	Research on features of CE	104
2.7	Objectives and research questions	113
2.8	Summary of chapter	116
Chapter 3:	Methodology and theoretical rationale	117
3.1	Introduction	117
3.2	Collocation theory	117
3.3	Research methodology: a corpus-based approach	126
3.3.1	What is a corpus?	126
3.3.2	Why use a computer to study language	128



Chapter 4: Research design	133
4.1 Research framework	133
4.2 Data	135
4.2.1 CEC	135
4.2.2 BNC	142
4.3 Preparation	143
4.4 Procedures	144
 Chapter 5: Results: CE collocational features	151
5.1 Research focus: Selection of items for analysis	152
5.1.1 Procedural steps	152
5.1.2 Node word, concordance and KWIC	154
5.1.3 Tokens, types, lemmas: Definition of terms	156
5.1.4 High frequency nouns in inner circle English	157
5.1.5 Narrowing down targets	164
5.2 Results: CE collocational features	169
5.2.1 Collocations	170
5.2.1.1 Overused collocations	171
5.2.1.2 Innovative collocations	190
5.2.2 Colligation	211
5.2.2.1 Enumerators as premodifiers in noun phrases	213
5.2.2.2 A preference for noun premodifiers	216
5.2.2.3 Genitive phrases as premodifiers of nouns	219
5.2.2.4 Postmodifying prepositional phrases	224
5.2.2.5 Postmodifying relative clauses	225
5.2.2.6 Conclusion of colligation features	227
5.2.3 Semantic preference	228
5.2.3.1 Selection of research focus	228



5.2.3.2	General observation of semantic preference	230
5.2.3.3	Semantic preference: PEOPLE, DEVELOPMENT	231
5.2.3.4	Conclusion of semantic preference	234
5.2.4	Semantic prosody	236
5.2.4.1	Selection of research focus	237
5.2.4.2	General observations of semantic prosody	238
5.2.4.3	Semantic prosody: PEOPLE, CHILD, SPIRIT	240
5.2.4.4	Conclusion of semantic prosody	252
5.3	Summary of chapter	253
Chapter 6:	Discussion	254
6.1	Existence of collocational features of CE	255
6.2	Word order in CE collocations	263
6.3	Rhetorical features	274
6.4	Collocational features of CE: A theoretical framework	280
6.5	Summary of chapter	284
Chapter 7:	Conclusion and implications	285
7.1	Recapitulation	285
7.2	Implications	289
7.3	Significance and values	293
7.4	From lexical variations to collocational variations	307
7.5	Limitations	314
7.6	Future research	314
7.7	Coda	315
References		319
Appendixes		341



List of Abbreviations

ABC	Australian Broadcasting Company
AmE	American English
BBC	British Broadcasting Company
BNC	British National Corpus
BrE	British English
CE	China English
CEC	China English Corpus
CET4	College English Test, band 4
EA	Error Analysis
ELF	English as a Lingua Franca
ELT	English Language Teaching
ESP	English for Specific Purposes
FLT	Foreign Language Teaching
GA	General American
HKE	Hong Kong English
ICE	International Corpus of English
ICT	Information and Communication Technology
RP	Received Pronunciation
SE	Singapore English
SLA	Second Language Acquisition
TESOL	Teaching English to Speakers of Other Languages
WE	World Englishes



List of Tables

Table 2.1:	Developmental cycles of new varieties of English	38
Table 2.2:	Proposed terms in reference to the English used by Chinese	46
Table 2.3	Speakers of English at the start of the 21 st century	64
Table 2.4	Distribution of proficient speakers of English and Mandarin worldwide	65
Table 2.5	The English learner populations in six major cities in China	66
Table 2.6	Two types of teachers in China	70
Table 2.7	Contrasting SLA with WE notions on CE	86
Table 2.8	Roles and status of English education in China	88
Table 2.9	Features of SE	107
Table 3.1	Four parameters of meaning	122
Table 3.2	Examples of popular corpora and their size	129
Table 3.3	Popular corpus tools	130
Table 4.1	Target corpus and reference corpus	135
Table 4.2	Domains of CEC and BNC	137
Table 4.3	Domains and description of the CEC data	139
Table 4.4	Contrasting procedures	146
Table 5.1	Procedural steps of identification of patterns	153
Table 5.2	Frequencies of word classes in written American and British English	158
Table 5.3	The 37 Lemmas shared by BNC and COCA (top 50)	160



Table 5.4	Target node words chosen	166
Table 5.5	AntConc operations	168
Table 5.6	20 overused collocations in CE	171
Table 5.7	CE collocations related to sociopolitical ideology in China	173
Table 5.8	Collocations related to culturally conditioned stance	177
Table 5.9	Linguistic structure feature	182
Table 5.10	Top ten nouns modified by <i>western</i> in CEC	186
Table 5.11	Top ten nouns modified by <i>western</i> in BNC	186
Table 5.12	CE collocations with <i>western</i>	187
Table 5.13	Top ten nouns modified by <i>foreign</i> in CEC	188
Table 5.14	Top ten nouns modified by <i>foreign</i> in BNC	188
Table 5.15	CE collocations with <i>foreign</i>	189
Table 5.16	Hyphenated compound adjectives as modifiers in CE	191
Table 5.17	Contrasting collocation <i>African-American people</i> in AmE and CE	192
Table 5.18	Innovative collocations in CE	197
Table 5.19	Contrasting the collocations of <i>all-round</i> in CEC and BNC	198
Table 5.20	More examples of innovative CE collocations	209
Table 5.21	Possible premodifiers of noun phrases	212
Table 5.22	Possible postmodifiers of noun phrases	212
Table 5.23	Examples of enumerators as premodifiers in CE noun phrases	213
Table 5.24	Types of nominal compounds in Mandarin	217



Table 5.25	Examples of using <i>people's</i> as a premodifier only in CEC	220
Table 5.26	Contrasting two-word clusters <i>people's</i> + <i>N</i> in CEC and BNC	221
Table 5.27	Main verbs in different grammar patterns	229
Table 5.28	Contrasting Mv.s for target lemmas in CEC	230
Table 5.29	Top 30 Mv.s collocate with DEVELOPMENT in CEC	232
Table 5.30	Top 30 Mv.s collocate with DEVELOPMENT in BNC	233
Table 5.31	Contrasting Adj.s for target lemmas in CEC and BNC	239
Table 5.32	Contrasting the adjectives pre-modifying CHILD in CEC	243
Table 5.33	Contrasting the frequency of occurrence of <i>individual CHILD</i>	243
Table 5.34	Contents of <i>Rules and Disciplines for Children</i>	245
Table 5.35	Common Chinese collocations of SPIRIT and 精神 (jīngshén)	248
Table 5.36	Contrasting Adj. pre-modifying SPIRIT	249
Table 6.1	Comparison of Chinese and English word orders	264
Table 6.2:	Summary of ratios of pre-modifiers to post-modifiers of WAY, FACT, THING in CEC and BNC	271
Table 7.1	Developmental cycle of new varieties of English	303
Table 7.2	Examples of CE lexis	308



List of Figures

Figure 2.1	English used around the world as native or official language	14
Figure 2.2	Graphic representation of the “Three Circles” model of English Speakers	19
Figure 2.3	Number of Chinese people studying abroad (2001-2013)	67
Figure 2.4	The statistics of students studying in United States (2007-2012)	68
Figure 2.5	The statistics of students studying in United Kingdom (2007-2012)	68
Figure 4.1	Research framework	134
Figure 4.2	Schematic procedure for the pilot phase of this study	145
Figure 5.1	KWIC analysis as exemplified by the keyword <i>revolutionary</i>	155
Figure 5.2	A sample wordlist based on one news headline	163
Figure 5.3	Comparative frequency data for target node words	167
Figure 5.4	From learner language analysis to language teaching	170
Figure 5.5	Contrasting frequencies of <i>common PEOPLE</i> in CEC and BNC	176
Figure 5.6	Contrasting frequencies of 人民 (<i>rénmín</i>) and 人们 (<i>rénmen</i>)	176
Figure 5.7	Contrasting the frequencies of <i>scientific way</i> in CEC and BNC	179
Figure 5.8	Contrasting frequencies of <i>scientific</i> in CEC and BNC	180
Figure 5.9	KWIC of <i>a scientific outlook on development</i> in CEC	181
Figure 5.10	Contrasting frequencies of <i>all-round</i> in CEC and BNC	183
Figure 5.11	Contrasting frequencies of <i>western PEOPLE</i> in CEC and BNC	184
Figure 5.12	Contrasting frequencies of <i>foreign PEOPLE</i> in CEC and BNC	185

Figure 5.13	Concordance of <i>plain MAN and WOMEN</i> in CEC	205
Figure 5.14	Concordance of <i>true FACT</i> in CE	207
Figure 5.15	Contrasting the proportions of nouns and adjectives as premodifiers	216
Figure 5.16	Examples of the pattern <i>people's + N of</i> in CEC	223
Figure 5.17	Contrasting the proportion of colligation <i>N's + N of</i> in CEC and BNC	223
Figure 5.18	Contrasting colligation <i>N + of</i>	225
Figure 5.19	Contrasting colligation <i>N + that</i>	225
Figure 5.20	INCREASE, SAVE, IMPROVE in L2-R2 position of PEOPLE	231
Figure 5.21	Contrasting semantic preferences in colour chart	234
Figure 5.22	Contrasting <i>common PEOPLE</i> and <i>ordinary PEOPLE</i>	241
Figure 6.1	Sample KWIC of <i>all-round WAY</i> in CEC	256
Figure 6.2	Sample KWIC of <i>eco-economic SYSTEM</i> in CEC	257
Figure 6.3	Sample KWIC of <i>Shanghai MAN</i> in CEC	258
Figure 6.4	Contrasting the numbers of pre-modifiers and post-modifiers of <i>WAY</i> and <i>FACT</i> in CEC	270
Figure 6.5	Contrasting the numbers of pre-modifiers and post-modifiers of <i>WAY</i> and <i>FACT</i> in BNC	270
Figure 6.6	Sample KWIC of <i>innovative spirit</i> in CEC	279
Figure 6.7	A proposed framework for analysing collocational features in CE	282
Figure 7.1	KWIC of <i>new INNOVATION</i> in BNC	310
Figure 7.2	KWIC of <i>new INNOVATION</i> in CEC	311



Figure 7.3	KWIC of <i>family RELATIVE</i> in BNC	312
Figure 7.4	KWIC of <i>family RELATIVE</i> in CEC	312
Figure 7.5	Doing and seeing things the Chinese and western way	316



CHAPTER 1

INTRODUCTION

The past decades have seen an unprecedented growth in the modes and diversity of language contact, partly due to forces of globalisation, and partly due to increasingly convenient access to the internet via all kinds of user-friendly smartphones and other mobile electronic communication devices (e.g., Blommaert, 2010; Crystal, 2001, among others). Mobile phones connected to the internet have largely redefined patterns of human interaction as well as modes of communication, which in turn help accelerate the processes of globalisation. Under these circumstances, diversity and change in English language usages should also be recognised, because “it is just as likely that the course of the English language is going to be influenced by those who speak it as a non-native language as by those who speak it as a mother tongue” (Crystal, 2012, p. 167).

1.1 Globalisation and the global spread of English

According to recent statistics published by International Telecommunication Union (ITU), a United Nations’ specialized agency for information and



communication technologies, in 2013, there were over 2.7 billion people using the internet, which corresponded with approximately 39% of the world's population (The World in 2013: ICT Facts and Figures, 2014, p. 2). Advancement in information and communication technologies (ICT) mediated by computers and smart phones and the increasing popularity of the internet have dramatically facilitated communication over a distance connecting people from different parts of the world.

Such forces of globalisation and the convenient access to a wide range of information via the internet have greatly favored the development of certain languages as lingua francas, especially English (Haberland, 2013). In an age of digital information, knowledge of English is required for gaining access to 'breaking news' broadcast in English. For instance, in the wake of the Edward Snowden news story from June, 2013, or the two doomed Malaysian Airlines aircraft MH370 (March, 2014) and MH17 (July, 2014), first-hand information tended to be made available in English. Internet users who wish to quickly find out and are keen to keep up with such news stories would have no access to such information if they did not know English. In addition, knowledge of English in following international news provides bilingual users of English with multiple perspectives as seen through different TV news agencies across different vantage points and contexts. The will to participate in events that



attract individuals' attention worldwide has reached a record high level. For example, after the 1-7 Brazil vs. Germany match on July 9th, 2014, some 66 million people around the world created (i.e. posted, discussed, or gave comments on) a total of over 0.2 billion postings on the Facebook pages 4 hours after the match (Retrieved from www.cnet.com). The active and instant mode of communication online has actually provided a strong incentive for bilingual users of English to get involved in meaning-making activities using new technologies.

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

Over half a century ago, in his book *Bilingualism as a World Problem*, Mackey (1967) proposed that the key reason for educated people to learn English is “to be able to seek the knowledge they need in one or more of the languages in which most of the world’s knowledge is available” (p. 18). Following dramatic advancement in ICT in the last few decades, however, affected are no longer



just the educated elite; rather, blue-collar workers in non-English-speaking countries who wish to be informed of and connected to other people may perceive a strong need to use English to understand what is going on elsewhere in the world and to communicate with people who do not share their first or usual language. Whatever their first language, educated or otherwise, people all over the world are connected; they are willing and readily facilitated to surf the internet to consume a wide range of information, from browsing breaking news and funny video-clips to searching for options of entertainment and leisure activities such as popular restaurants and internet games with the most ‘Likes’. The internet also provides a platform for social interaction, such as chatting with friends at e-forums, sharing opinions in blogs and giving ‘Likes’ on Facebook pages. Such out-of-class, internet-mediated activities have become an integral part of our everyday life; the wide-ranging needs for information in English plus a strong motivation to be connected with people from different language backgrounds has given English a special status, a lingua franca for international communication. This need for communication continues to maintain English as the language that people cannot afford not to learn and use nowadays.

Commonly referred to as ‘the world’s lingua franca’, English plays a vital role in the globalisation-in-a-multilingual-world movement, which results in an



“English language complex” (Mesthrie & Bhatt, 2008, pp. 1-3). “We’re all ‘global’ now, and need to use the first truly universal language, whether we are business people, politicians, teachers, tourists, or terrorists” (McArthur, 2003, p. 54). “What happens, linguistically, when the numbers of the human race use a technology enabling any of them to be in routine contact with anyone else?” (Crystal, 2001, p. 5).

For historical, social and ICT-related reasons, it is already a fact that the spread of English around the world has been, and continues to be both rapid and unpredictable. One consequence of colonisation (in the past) and globalisation (at present) is the general spread of English on one hand, and inclusion of English language teaching in the national or local education curriculum on the other (Jenkins, 2009). Being an increasingly important and active player in global political and economic affairs, China, with the largest numbers of users of lingua franca English in the world, is one such nation where the above-mentioned forces of globalisation and growing popularity of the internet are played out. Users of English in China face many challenges. These include: a) English has become an important part of compulsory education, lasting for 13 to 15 years from primary to tertiary education depending on the onset year (Grade 3 or Grade 1); and b) the objective need as well as subjective wish to be increasingly engaged in the English-dominant internet makes it necessary for



millions of Chinese nationals to use English more or less regularly, both as consumers of information and as agents interacting with people from different language backgrounds. Such conditions have already greatly influenced the language ecologies in China today, especially the role it plays for the nation as well as for individual citizens. Politically, English has become one very important tool for the nation to build up her image and articulate her diplomatic stance in international, especially bilateral issues. Domestically, multilingual users of English now enjoy more and more global horizons via English. In short, the sociopolitical role of China in the world, together with the rapidly increasing use of English-oriented information and communications technologies constitutes the background and setting for Chinese people to use English as part of their daily-life activities.

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



English as used in Mainland China is the variety I will concentrate on in this thesis. More specifically, our main focus is on the features of English as found in the written outputs gathered from Han Chinese authors in Mainland China. A historical overview of the factors and circumstances that gave rise to them will be provided. We will first review debates about China English (CE), the possibilities of codifying CE patterns and other recent developments from a World Englishes perspective. Then we will present and describe a detailed account of a number of salient structures and collocation patterns. This will set the scene for the proposed theoretical framework for uncovering CE collocational features.

The English as used in the two Special Administrative Regions (SARs), Hong Kong and Macau, will not be covered in this thesis. There are two main reasons for this. Firstly, the sociolinguistic backgrounds to the use of English in the two SARs are very different. Historically speaking, English in the Mainland, Hong Kong and Macau had different functions and statuses and followed rather dissimilar paths of development. Socio-politically, there is some similarity between Hong Kong and Macau: both have had a lengthy history of colonisation, as a result of which the language environment as well as western culture have rooted deeply in these two communities. The main difference is that Hong Kong SAR was colonised by Great Britain, which resulted in a



longer English-using history as well as an official status for English, which is why English has traditionally played an important institutional role in Hong Kongers' lifeworld. Macau, on the other hand, was colonised by the Portuguese and "throughout much of the colonial period, Portuguese was the only official language" (Leung & D. C. S. Li, 2011, p. 86). As such, the use of English in Macau was understandably not the same as that of Hong Kong and the Mainland, where no other official language was ever used in her history except Mandarin Chinese.

Secondly, research shows that Hong Kong English and Macau English have evolved in different stages. The status of Hong Kong English has evolved from an 'auxiliary' to a 'value-added' position since post-1997 Hong Kong (D. C. S. Li, 1999, p. 67). Bolton (2003) believes that "essential conditions necessary for the emergence of such a variety may already be present in the community" (p. 116). The recognition of Hong Kong English as an emerging variety of English is increasingly accepted (cf. Bolton, 2000, 2002, 2003; Evans, 2011; Kirkpatrick, 2007; D. C. S. Li, 2007, 2010a; Sewell, 2009). "Hong Kong English will eventually be pushed more firmly towards Kachru's outer circle, Schneider's phase 4" (Setter, Wong & Chan, 2010, p. 116). Macau English, unlike Hong Kong English, has been "overlooked as a variety" (Moody, 2008, p. 14). Signs of autonomous use of English can be detected. For instance, Young



(2006) found that students in Macau have “strong motivation to learn English and readiness to use English as a medium of instruction” (p. 479); “70% government institutions that offer services on the internet offer them in English in addition to Chinese and Portuguese” (Moody, 2008, p. 4). However, the use of English is not as frequent as in the case of Hong Kong. Botha (2013), for example, reported that English as a medium of teaching and learning in higher education is promoted but not actually implemented as expected (p. 475). So “complete exposure to English” across the Macau community is not the norm (Botha, 2013, p. 474), partly because “the colonial administration seemed laissez-faire or indifferent as to their language policy in Macau” (cited in Leung & Li, 2011, p. 86). However, features of Macau English can still be detected, for example, “high-rising intonation is typically used in questions” and words adopted from Cantonese “*laissee, bat gwa*” (Moody, 2008, pp. 10-12). To conclude, Hong Kong and Macau SARs are different sociopolitical entities that deserve separate attention and study in terms of the development of English as a localized variety. It is for these reasons that they are not included in this thesis.

1.2 Organization of the thesis

This thesis consists of seven chapters. Chapter 1 will briefly outline the background of the present study. This will be followed by a literature review of



research on WE and CE in the second chapter. Chapter 3 will provide an overview of the theoretical rationale, where a corpus-based perspective in collocation studies of English varieties will be discussed. Chapter 4 will present the research design adopted in this study, including the choice of methods and types of data. Chapter 5 will outline the analytical framework, report the results of the analysis, and how these address the research questions. Collocations of high-frequency nouns in CEC (China English Corpus) will be studied and analysed. Patterns ranging from specific collocations to more abstract domains like colligations and semantic/prosodic preferences will be clarified and summed up. Chapter 6 will discuss the significance of the findings from different perspectives, with a view to making generalisations and presenting an overview of the key findings. Chapter 7 will conclude this study by offering a critical account of the implications of its key findings, and demonstrating their connections to the research questions as stated in section 2.7.



CHAPTER 2

LITERATURE REVIEW

As modern technologies such as the internet and ‘apps’ of all kinds increasingly facilitate communication and interaction between people, language contact phenomena become increasingly attractive to linguists. Meanwhile, as a result of education, ‘bilingual speaker of English’ has become a common attribute or characteristic for the majority of people all over the world, regardless of their first language background. The interface of these two trends is English, a focus of concern to critical sociolinguists. It is quite possible that English will continue to play a role in China’s engagement with the world, and under these circumstances, the very large population of CE users is likely to attract more and more researchers to investigate issues arising in this field.

A ‘bond’ between China and English language could be detected. Thus this part of the thesis aims at situating CE in the global context of wider discussions in the closely related research paradigms of World Englishes (WE) and English as a lingua franca (ELF), and provides a brief historical sketch of research undertaken in these fields.



2.1 The World Englishes perspective

Braj Kachru's pioneering 'three concentric circles' model (1982, 1985, 1986, 1990, 1996, 2003), representing the growth and spread of English in the world, has been invaluable for researchers in the field of WE to come to grips with the implications of the pervasive, highly complex and infiltrating presence of English in this second decade of the twenty-first century. While the 'three concentric circles' model is not the only theoretical account of the tremendous variation and spread of English worldwide, it is certainly the most influential, and its impact may be felt and tracked in research on WE in the last three decades and, more recently, English as a lingua franca (ELF). Englishes in different countries and regions are increasingly regarded as "pluricentric languages" (Clyne, 1992), a view which is perfectly compatible with the WE perspective.

2.1.1 Models of Englishes

Descriptions of the classifications or models of English language have been proposed by a number of scholars who attempt to explain the differences in the ways English is used in different localities.



2.1.1.1 ENL/ESL/EFL distinctions

One of the early attempts to account for systematic variation in the way English is used by speakers in different parts of the world is a tripartite distinction between English as a Native Language (ENL), English as a Second Language (ESL), and English as a Foreign Language (EFL) (Quirk, Greenbaum, Leech & Svartvik, 1985; Conrad, 1996). Based on this classification, Stern (1983), Widdowson (1997, 2003), McArthur (1998), Nunan & Carter (2001), among many other scholars, have proposed definitions and discussed the importance of these distinctions. Some other linguists (see, e.g., Görlach, 2002, pp. 99-117), have also tried to tackle theoretical issues arising from the use of these terms from a more or less WE perspective.

Crystal (2012) provides a picture to help visualise the way English is used as a native language around the world at the time of writing (Figure 2.1). Countries in red are places where English is used as mother tongue or native language. Countries in purple represent places where English is used as an official language. The places in green, as proposed by Crystal, are places where English is used as a mother tongue, but co-exists with other languages. He does not explain whether it is implied that the places besides these three coloured zones (which obviously constitute a larger territory compared to the above-mentioned)



There is no doubt that this ‘ENL/ESL/EFL distinction’ has been helpful in certain contexts, but it has its shortcomings. One obvious drawback is that the notion *native* versus *second* and *foreign* indicates considerable linguistic prejudice against the latter terms, which might imply that one is superior and the others should look up to it as a standard. Another disadvantage may be that it is increasingly difficult to find countries that can be accurately classified as using EFL. How can one possibly distinguish ‘second’ from ‘foreign’ in the 21st century, with the internet, Wi-Fi, Facebook, and built-in smartphone functions such as WhatsApp creating such readily accessible communication platforms connecting people regardless of the physical distance between them? For example, when surfing the internet and using communication tools like Facebook, people might use English most of the time. How can we define one user of this kind in Singapore as an ESL user while another in China as an EFL user?

Consider the case of one ordinary child in Guangzhou City, Mainland China, where he/she usually learns to speak his/her mother tongue (Cantonese) around one year old. And immediately after that, he/she would most probably be sent to Early Education Center/Early Childhood Development Academy (早教中心) for pre-primary tutoring activities, for example *Disney English* (<http://disneyenglish.com>), *iKid English* (<http://www.ikidenglish.com/>), *iBaby*



English (<http://edu.pcbaby.com.cn/center/iBabyEnglish>), etc., where English is taught, with Mandarin as the medium of instruction. In such cases, English and Mandarin are acquired quasi-simultaneously, making it difficult to assign the term EFL and ESL to young people growing up with such experiences. The term ENL is another problematic label in a setting like Mainland China. Take the case of ‘foreign experts’ teaching English in China. It matters more to students, and their parents, that the foreign experts are proficient in English, regardless of whether they are ENL speakers from UK, USA, Canada, Australia or New Zealand (for more details, see Section 2.4.2). Usually the tutors in the language training schools/centers come from all over the world: Singapore, Europe and Africa, for example. The term ‘native’ is not as important as the fact that they can speak fluent English, which caters to the needs of the parents who pay for the courses.

These examples show how futile (and sometimes meaningless) the ENL/ESL/EFL distinction is to the general public, even though such a three-fold distinction or classification is well motivated theoretically speaking in research on the spread of English worldwide. So in the field of ELT, such a tripartite distinction is still widely acknowledged and held to be important by many scholars.



2.1.1.2 Inner/outer/expanding circles

An alternative and equally influential classification of English, the inner/outer/expanding circles distinction, has been put forward by Kachru (1982) and supported by many others. One of the grounds for this classification is:

The situation of English is historically and linguistically interesting and complex. First, the number of non-native speakers of English is significant; second, the spread of English is unique. The large range of varieties of English cannot be discussed from any one point of view. There are several, mutually non-exclusive ways to discuss their form and function. (Kachru, 1982, pp. 36-37)

Kachru has been advocating this view since the mid-1960s, when he pointed to the ‘Indianness’ of Indian English (1965, p. 391). His contribution attracted considerable scholarly attention in the 1970s, when two important international conferences were held which independently focused on a shared theme: the features and functions of the English varieties used in non-native speaking countries. One was organised at the East-West Culture Learning Institute of the East-West Center, Hawaii, and the other at the University of Illinois at Urbana-Champaign, resulting in the publications *English for Cross-cultural*



Communication (Smith, 1981) and *The Other Tongue: English across Cultures* (Kachru, 1982), respectively.

Since then, the study of varieties of English, known as World Englishes, has become “a new sub-discipline of English linguistics” (Schneider, 2003, p. 234). The plural form ‘Englishes’ symbolises “variation in form and function, use in linguistically and culturally distinct contexts, and a range of variety in literary creativity...it stresses the ‘WE-ness’ among the users of English, as opposed to *us* versus *them* (native and non-native)” (Kachru, 1996, p. 135). In order to have a better understanding of the use of English in different countries, Kachru (1985) put forward a three-concentric-circles model of World Englishes:

The current sociolinguistic profile of English may be viewed in terms of three concentric circles... The inner circle refers to the traditional cultural and linguistic bases of English. The outer circle represents the institutionalised non-native varieties (ESL) in the regions that have passed through extended periods of colonisation...The expanding circle includes the regions where the performance varieties of the language are used essentially in EFL contexts. (pp. 366-367)

For decades, this has been considered an influential model that accounts for the



processes as well as the outcomes of the spread of English to date. There are others, however, including Streven's *World Map of English* (1980), McArthur's *Circle of World English* (1987), and Görlach's *Circle Model of English* (1990). In this thesis, we will follow Kachru's model, where the inner circle refers to the traditional cultural and linguistic bases of English, namely the USA and the UK, Canada, Australia and New Zealand. The outer circle includes countries where English is not the native tongue, but is important and plays a part in the nation's institutions, either as an official or national language (e.g., India, Nigeria, Singapore, the Philippines, Pakistan). The expanding circle includes the regions where the performance varieties of the language are used essentially in lingua franca contexts, for example, China, Egypt, Japan and Korea.

Though it is not possible to calculate populations of speakers of the three circles, estimates have been attempted in different ways. One example of this is from Crystal (1997):

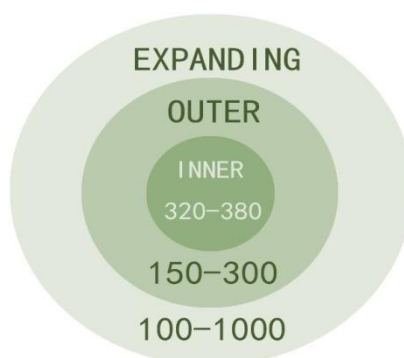


Figure 2.2: Graphic representation of the “Three Circles” model of English Speakers (in millions) (adapted from Crystal, 1997, p. 54)

These numbers of speakers are largely underestimations now, because they were cited over 15 years ago. Reliable updated numbers for the three circles of speakers are not easily obtainable. Some reasons are rather obvious. For the number of speakers in the inner circle, the differences of options as to which countries/regions should be included under that heading will probably be a key reason that hinders the counting. The total number of expanding circle speakers are even more difficult to be sure about, for the obvious reason that fluency is a continuum whereby people will differ in their view about how much competence of English a person needs before being counted as a member of the community of expanding circle users. This is probably why the numbers in Figure 2.2 are still widely cited even today.

The evolving nature of the English language nowadays has called for a reassessment of a number of key dimensions in the analysis of English varieties. The foregoing discussion in regard to the question ‘what form English should (or is likely to) take as a model to follow’ has become a naïve yet crucial question to answer. Which groups are empowered and which ones disadvantaged? Why? Should we or should we not accept that fact, namely that “People from the so-called ‘core’ English speaking countries are now in the minority among English users and ‘native-speakers’ of the language no longer determine how the language is being used internationally” (Clyne & Sharifian,



2008, p. 281)? Would people from the inner circle countries feel upset by statements such as “English is now a new phenomenon, and if it represents any kinds of triumph it is probably not a cause for celebration by native speakers” (Graddol, 2006, p. 11)? These are among those questions that we can start asking, when we accept a WE perspective toward today’s Englishes.

2.1.1.3 Global language system: peripheral/central/supercentral/hypercentral languages

Besides its global significance in terms of its functions and status, English is also thought to play a key role in the “global language system” (de Swaan, 2001, pp. 1-4). According to the “global language system” theory, the world’s languages are hierarchically connected as one system, in that they fall into a pyramid-like hierarchy constructed by four levels. They are named as the peripheral, central, supercentral and hypercentral languages, with peripheral languages being situated at the lowest level of the hierarchy, serving principally as “languages of conversation and narration rather than reading and writing, of memory and remembrance rather than record” (p. 4). According to de Swaan, 98% of the world’s languages are peripheral languages used by less than 10% of people in the world today. It is foreseeable that, being used only by speakers in particular areas (e.g., Jun dialect in China is estimated as spoken by only



thousands of people in remote areas around South-east China) of the world, many peripheral languages are at risk of falling out of use and facing the danger of extinct. At the same time, with increasing globalisation, more and more speakers of peripheral languages acquire more central languages in order to communicate with others. The next level — central languages — is constituted by 100 languages, which are spoken by about 95% of the world's population and generally used in the domains of education, public media and administration, and which typically serve as the national/official language(s) of the ruling state. Central languages are “languages of record: much of what has been said and written in those languages is saved in newspaper reports, minutes and proceedings, stored in archives, included in history books, collections of the ‘classics’, of folk talks and folk ways, increasingly recorded on electronic media, and thus conserved for posterity” (ibid, p. 5). Taking the society as a whole, many speakers of central languages become multilingual eventually. Because they are either native users of one particular peripheral language who have learned a central languages or native speakers of a central language and have acquired a supercentral language, the latter being situated at the second-highest level of the language hierarchy. Supercentral languages are usually widely-used languages that serve as connectors between users of central languages (i.e., Arabic, Chinese, English, French, German, Hindi, Japanese, Malay, Portuguese, Russian, Spanish, Swahili and Turkish, de Swaan, 2001, p.



5). Colonial traces for these languages could usually be observed. In other words, they “were once imposed by a colonial power and after independence continued to be used in politics, administration, law, big business, technology and higher education” (ibid, p. 5). The ‘world language system’ theory proposes that, at the upper-most level is the language which connects speakers of the supercentral languages — the hypercentral language which “holds the entire world language system together. Currently, this hypercentral language is English” (ibid, p. 17). As such, the world’s individual multilinguals are connected in a markedly hierarchically structured pattern. Though de Swaan’s theory of the world’s language system has been critiqued, with various scholars expressing concerns about its lack of a solid, empirically-grounded theoretical foundation (e.g., Morris, 2004; Phillipson, 2004), his conceptualization of a hierarchical pyramid of the languages of the world still represents an influential perspective of how natural languages are related in one world language system. At the same time, de Swaan provides a theoretical account on how English may link up individual multilinguals whose L1 is not English, thereby placing English in an (especially IT-mediated) functional position, which complements the communicative functions of English as seen from the conceptual lens of Kachru’s nations-based concentric circles of English.



2.1.1.4 Summary

The ‘ENL/ESL/EFL’ and the ‘inner/outer/expanding circles’ distinctions, as well as the ‘global language system’ theory are both important in their own ways, and neither should be judged intuitively to be more valid than the other. Obviously, some countries and speakers in the two classification models would inevitably overlap. They are by no means mutually exclusive; quite the contrary, insights obtained in one tend to enlighten or inform further research in the other.

2.1.2 Norms and standards

English has a long history and, far from being static, its standards have been gradually undergoing change over time. According to Svartvik and Leech (2006), English was a collection of dialects spoken by marauding, illiterate tribes some 1,500 years ago, so it evolved from the grassroots rather than from society’s elite. The question of standardness has always been a controversy, due to the complex evolution history, even with respect to the inner-circle varieties.

2.1.2.1 Standard Englishes

Standard English is the variety of English which is considered to be the norm



(Jenkins, 2003, p. 29). The term ‘standard’ usually indicates acceptability. Out of common sense, this standard should generally be provided by the ‘native’ speakers of English, codified according to given norms at various linguistic levels, which may be used to ascertain whether a learner has attained the level of proficiency required according to these norms. This assumption is based on the same logic as agreeing on the rules for a sports game.

Many scholars have attempted over the years to provide answers to the question of ‘What is standard English’, either positively or negatively (Jenkins, 2000, 2007; Cameron, 1995; Crowley, 2003; Trudgill & Hannah, 2008, Holmes, 2013). There is general agreement on the idea of ‘standard’ being driven by a ‘standard language ideology’, which may be described as:

A particular set of beliefs about language. Such beliefs are typically held by populations of economically developed nations where processes of standardisation have operated over a considerable time to produce an abstract set of norms—lexical, grammatical and...phonological—popularly described as constituting a standard language. (Milroy, L. 1999, p. 173)

Here, though Milroy is talking about ‘language’, English is definitely one of the key referents. As can be inferred, there are people who do not share such an



ideology, whether it is ‘language’ or ‘English’. So it is not surprising to find that some believe ‘standard English’ to be a false proposition, such as Ward (1929, cited in Kachru, 1982, p. 34), who states that “No one can adequately define it, because such a thing does not exist”. On the other hand, there are scholars who believe that Standard English does exist and offer their respective definitions, for example, Trudgill (1984), Strevens (1985), Crystal (1995), among many others.

One of the best known attempts to describe it is found in an article by Trudgill (1999), ‘*Standard English: what it isn’t*’. Trudgill suggests that the only feasible way of defining a ‘non-academy’ standard language such as English is to demonstrate what it is not (pp. 117-128):

Not:	But:
a language	one variety of given English
an accent	spoken by 12-15% of the population in Britain, of which 9-12% speaks it with a regional accent
a style	can be spoken in formal, neutral and informal styles
a register	no necessary connection between register and Standard English
a set of prescriptive rules	tolerate certain features which prescriptive grammarians do not allow

The point Trudgill (1999) makes is that Standard English is purely a social dialect, albeit a more prestigious one. Though there are certain differences between this and other dialects, it does not necessarily indicate the linguistic superiority of the standard forms. Thus exactly what constitutes standard English may be illusory and sometimes difficult to ascertain, especially in the process of learning English.

An example of my own experience as a student is relevant here: the collocation ‘teacher of English’ was the only version of choice in test papers in the 1980s, in contrast with ‘English teacher’. The reason for the non-idiomatic status of the latter collocation, as I could recall from my teacher’s lecture, was decided and followed because the textbooks of English in China said so. ‘English teacher’ was considered to have an ambiguity of referring to ‘teacher from England’. In Mainland China, ‘teacher of English’ was the standard form in the 1980s, while ‘English teacher’ was not. However, both expressions are considered synonymous and quite acceptable nowadays, as we could find 101 tokens of ‘English teacher’, and 63 tokens of ‘teacher of English’ in BNC. And more importantly, these standards are also followed in Mainland China today as teachers will no longer challenge the use of ‘English teacher’ in students’ writing now. This example illustrates that what counts as an established norm or standard in an inner circle variety (BrE) may be dismissed by users in the



expanding circle as unacceptable at one point, but accepted later.

In actual fact, there is never an organised agency vested with the mission and authority to provide the direction toward setting up a standardised model of usages for English, controlling or containing changes as a result of language contact. Such attempts did occur in the past, however (e.g., Jonathan Swift in Britain proposed one around 1712, and John Adams in 1821, although neither was received with enthusiasm; cited in Heath, 1977). What factors have determined the correctness of the linguistic ‘etiquette’ in English then? Kachru (1983) believes that the “sanctity of models of English stems more from social and attitudinal factors than from reasons of authority. These models, more widely violated than followed, stand more for elitism than for authority—and in this sense they have a disadvantage” (p. 32). If we consider this issue from a diachronic point of view, the standards and norms of English have shifted considerably over time. Elitist prescriptivism can be found both in history and today; they only differ in degree. Puttenham wrote in his book *Arte of English poesie* published in 1589, that “the model of standard English should be the usual speech of the court, and that of London and the shires lying about London within 60 miles and not much above” (cited in Kachru, 1982, p. 33). In the 17th century, similar statements single out the speech of London or that of the universities of Oxford and Cambridge as the best. Conversely, regional dialects



receive adverse comments and are regularly used to characterize naive or rustic speakers. Henry Sweet (1910, p. 7) wrote: “Standard English...is the language of the educated all over Great Britain”. Later on, non-native speakers of English aimed at a close approximation of two well-documented models, Received Pronunciation (RP), and General American (GA), as models of pronunciation to follow as well as the yardstick against which other varieties of the language are to be measured. However, “although speakers believe in the existence of some canonical form, the language continues to vary and change” (Milroy, J. 1999, p.17). Another key player for this argument is Bamgbose (1998), who makes the following prediction:

Today, few serious scholars of the English language will insist that a non-native English is used only in a narrow range of domains, that it is a transitional and unstable code striving for perfection, that its continued encouragement and use will lead to linguistic fragmentation and/or deterioration of the English language, or that only native English is a suitable model for all English language users. (1998, p. 1)

The advent of British Broadcasting Corporation (BBC) played an important role in making RP widely known, a prestigious accent that was subsequently termed “BBC English” (Gimson, 1970, p. 83). However, assigning the value of such



norms for teachers and learners of English has not been universally accepted.

Abercrombie (1965) made three points to illustrate his objection to this:

First, it [conferring social prestige to RP] is a bad thing rather than a good thing. It is an anachronism in present-day democratic society. Second, it provides an accent-bar which does not reflect the social reality of England. Finally, while those who talk RP can justly consider themselves educated, they are outnumbered these days by the undoubtedly educated people who do not talk RP. (p. 15)

As far as GA is concerned, Kenyon suggested in 1924 that “it refers to the variety of English spoken by about 90 million people in the central and western United States and in most of Canada” (cited in Kachru, 1982, p. 34). The difference is that Kenyon also advocated linguistic tolerance toward various American varieties of English; he was concerned that “we accept rules of pronunciation as authoritative without inquiry into either the validity of the rules or the fitness of their authors to promulgate them” (p. 34). He suggested that the cause for such easy judgement connected with pronunciation is that people are just “influenced by certain types of teaching in the schools, by the indiscriminating use of textbooks on grammar and rhetoric, by unintelligent use of the dictionary, by manuals of correct English, each with its favorite



shibboleth” (p. 34). This argument seems to be common sense. “Probably no intelligent person actually expects ‘cultivated’ people in the South, the East, and the West, to pronounce alike” (p. 34). In line with this argument against the use of RP or GA model, Jenkins (2000, 2007) has put forward other compelling arguments against the hegemony of RP/GA model : “Whatever ‘circle’ we come from, we all need to think about why we make our linguistic choices and what attitudes and beliefs inform the identities we accept for ourselves and ascribe to others” (2007, p. 233).

The implication behind all this is that believing in and adhering to some ‘pure standard’ is not such an important matter to be taken seriously by language users, for standard itself is not as relevant as other factors such as users’ attitudes and identities associated with English. Of course, the question of whether one needs to adhere to ‘a standard English variety’ or not is far from simple. One related issue is linguistic prejudice. Two enquiries into linguistic prejudice are appropriate here, both by Giles & Powesland (1975). The first study aims at investigating British school children’s reaction to a variety of English accents, namely RP, the Birmingham accent, and a number of other rural and urban accents. Participants were asked to listen to speakers with the above accents, and make judgements about them. Results show that people who spoke with a standard RP accent were considered the most intelligent, and those

with a Birmingham accent the least intelligent. Another finding was that participants felt that the speakers with accents closest to their own were more honest and warmer people. As a follow-up study, Giles and Powesland designed another project investigating people's responses to arguments about capital punishment, given in written transcript form, and read in four different British accents, namely RP, Welsh, Somerset and Birmingham. Children who participated in the research reported that arguments given in RP were more intelligent; however, they also reported that arguments in regional accents were more persuasive because they "were effective in producing a significant shift in subjects' attitudes; the typescript and the RP guise did not" (Giles & Powesland, 1975, p. 94). While attempts are made to set RP and GA as norms for pronunciations of Standard English for the world to follow, disagreements and challenges such as those cited in the earlier part of this chapter should not be taken lightly. Norms of this kind, no matter how appropriate, are sometimes difficult if not impossible to follow. This may not present a serious problem to ordinary people, but it will surely be an issue for language teaching curriculum writers. For example, among the six 'National Syllabus for English Teaching' in China (after the *Cultural Revolution*, seven versions were issued in 1978, 1981, 1986, 1992, 1993, 1996, 2003), the standards for teaching and learning consistently remain vague. The word 'standard' is set as goal but it is largely a myth because problems of standardness remain.



Similar problems are faced in Hong Kong with its history of being a former colony and over a hundred years of experience of contact with the English language. A place which was ruled by the British seems to continue struggling against the problem of norms, a problem associated with Hong Kong's language policy of biliteracy and trilingualism: "a high level of English proficiency...is likely to remain a remote if not unattainable goal" (D. C. S. Li, 2010, p. 109). Not only Hong Kong; it seems likely that English teaching in China will also continue to face a similar problem. Perhaps the most successful solution, from those so far proposed, would be a better and more realistic pedagogical norm for English language teaching. The debates surrounding this issue can be found in more detail, for example, in Jenkins (2006, 2007), Murata and Jenkins (2009), Seidlhofer (2004, 2006), McKay (2003, 2005, 2009), D. C. S. Li (2007), Graddol (2006), Bolton, Graddol and Mesthrie (2011).

To conclude, we now live in an era when multilingualism and multiculturalism should be viewed as the norm; when the trials and tribulations of newcomers should be received with tolerance and empathy; when deviations are not a problem, and when 'speaking in many tongues' is incorporated into mindsets of all citizens, such deviations from native-speaker standards are more appropriately seen as acts of identity (Le Page & Tabouret-Keller, 1985) than as violations of native-speaker norms and standards.



It should be noted, however, that the question of standardness sheds more important light on certain groups of people for whom English is something more than speech sounds associated with a standard variety. Apart from such influential institutions as formal education, grammar books and dictionaries, key players in the national media including news anchors and editors, educated users of English worldwide (regardless of their first-language background, i.e. including educated users of English in Kachru's outer circle and expanding circle countries) make up another important monitoring body. Their English is more likely to be looked upon as model to be followed, so there is a stronger need for standards in these fields. However, the justifiable need for a set of well-defined standards does little to weaken the critique of standard English outlined above, since the English used in such formal domains is arguably different forms of English for Specific Purposes (ESP), where there is clearly a need for institutionalised, more stringent standards compared with the English used for general communication purposes such as using English to exchange information or signal rapport in social interaction with others. ESP, "an approach to language teaching which is directed by specific and apparent reasons for learning" (Hutchinson & Waters, 1987, p. 19), is very much concerned about norms and standards by definition. Among the many types of ESP (e.g., English for economics, English for engineers, English for nursing (ibid., p. 17), teaching and learning activities plus assessments are all driven by



goals and standards (which cater to the needs of the specific purpose) set in the curricula and teaching materials. One of the purposes of learning the language in ESP is to communicate within a specific academic, occupational or professional domain. So standardness sheds more important light on ESP teaching and learning. To my knowledge, Hutchinson and Waters (1987), Robinson (1991), Jordan (1997), Dudley Evans and St. John (1998), and Gavioli (2005) are key players in this field.

Another important dimension of the issue of English standards concerns language use on the internet, where English has emerged as a *de facto* repository of normative usage. It seems that the emergence of the internet, especially when the world-wide-web is accessible to mobile phones users, has given rise to many concerns and frustrations about English as used on the internet. Do the relaxed standards of e-mails augur the end of literacy and spelling as we know it? Will the internet herald a new era of buzzwords? Will linguistic creativity and flexibility be lost as globalisation imposes uniformity of norms? These questions are among those widely held anxieties about the effect of the internet on language(s). As this issue is beyond the scope of inquiry in this study, we will not go into further details here (but see Crystal, 2001, 2012).



2.1.2.2 Non-standard Englishes

The antithesis of ‘standard English’ is ‘non-standard Englishes’, a label which has been assigned to all the New Englishes. There is much to be said about this topic, but what is most relevant here is the fact that non-standard Englishes are not static either, in much the same situation as changes within standard English varieties. So under non-standard Englishes, our main focus is on innovation and development via actual language use, rather than deficiency of language proficiency in the learning process.

Perhaps a best case that shows a non-standard English can get acceptance and become an inner circle variety is Australian English. According to Dixon (2011), there were some 250 Aboriginal languages in Australia when Captain Cook arrived in 1770. Yet the dominant place of English becomes stronger and stronger as time goes by. Some of the standard varieties today were historically regarded as non-standard previously. This is clearly reflected in the following two examples. Görlach (1991) quotes an 1829 source commenting on Australian English as follows:

Bearing in mind that our lowest class brought with it a peculiar language and is constantly supplied with fresh corruption, you will understand why



pure English is not, and is not likely to become, the language of the colony.

(p. 147)

Another example occurred in 1920, when the Director of Education in New South Wales urged teachers to stop using Australian English in the classroom by leaving a note saying “It is sad to reflect that other people are able to recognise Australians by their speech” (Delbridge, 1999, p. 260). This view was widely shared at that time, reflecting people’s perception of Australian English being non-standard.

Prejudice continued to be voiced against this ‘non-standard’ English until 1952, when the Australian Broadcasting Commission (ABC) started to accept Australian English in their programmes, after Arthur Mitchell, a linguist who promoted the idea that Australian English should be an acceptable norm, became the Chairman of ABC. Later on, the publication of the following three dictionaries helped Australian English to become established as a standard:

- a) *A Dictionary of Australian Colloquialisms* (Wilkes, 1978);
- b) *The Macquarie Dictionary* (Delbridge, 1981);
- c) *The Australian National Dictionary* (Ransom, 1988).



Nowadays, more dictionaries of Australian English have been published, yet not comparable to these three, in that they set out a milestone flagging a significant change in the status of Australian English: from non-standard variety to standard variety. The case of Australian English also conforms to Schneider's (2003, 2007) dynamic evolutionary theory of the emergence of new varieties of English, which is characterised by developmental cycles or phases (Table 2.1).

Table 2.1: Developmental cycles of new varieties of English (Schneider, 2003)

Phases	Description
1. Foundation	English begins to be used in a country, typically because English speakers settle in the country
2. Exonormative stabilisation	the variety spoken by the local people (IDG) is closely modelled on the variety imported by the settlers (STL)
3. Nativisation	the establishment of a new identity with the coupling of the imported STL and local IDG varieties; reconstruction of vocabulary and grammar occurs in IDG, which poses heavy effects on STL at the same time
4. Endonormative stabilisation	IDG becomes gradually accepted as the local norm or model and is used in a range of formal situations
5. Differentiation	IDG has emerged, reflecting local identity and culture

About a decade later, Schneider (2014) updated his theory towards the developmental cycles of new varieties of English with the following observation:

I suggest that the current rush towards and the multiple applications of Englishes on a global scale represent a process best conceptualised as ‘Transnational Attraction’—the appropriation of (components of) English(es) for whatever communication purposes at hand, unbounded by distinctions of norms, nations or varieties. (Schneider, 2014, p. 28)

2.1.2.3 Summary

“The very fact that English is an international language means that no nation can have custody over it” (Widdowson, 2003, p. 43). The case of Australian English shows that, on one hand, it is in the inner circle of English but not ‘naturally’ being so; on the other hand, it is accepted and institutionalised after undergoing evolution and development over an extended period. In the case of Australian English, it took 200 years to finish the cycle of development. So it may be argued that CE, an emerging variety and focus of this study, is subject to similar developmental procedures. However, for English language teaching, what matters most may be the teaching norms rather than the question ‘which



circle CE belongs to'. It seems clear that there are preferred ways of saying things under certain conditions, and these might stem from the British or American norms; however, as far as CE is concerned, there is no need to adhere strictly to a standard without considering cultural factors, especially in the era of the internet, when innovations and change in English (among other languages) are spread and taken up more quickly than ever before. If one accepts the fact that there is no real single standard for everybody in the world who speaks English to refer to as the norm, the Polymodel Approach should be used to replace the Monomodel Approach (Kachru, 1982, p. 66), given that English is not a single, homogenous, monolithic variety. Following this ideology, the variability of the English language should be recognised and highlighted. Furthermore, this is one of the advantages of the three-concentric-circle model over the ENL/ESL/EFL demarcation. After half a century of development, WE has evolved into a discipline following progressive refinement of its theoretical underpinnings, such as those influential monographs by Kachru (1982, 1985, 1990, 1992), Bolton and Kachru (2006), Kirkpatrick (2007, 2010), McArthur (1998, 2002), Schneider (1997, 2003, 2010), Smith (1992), Jenkins (2003), Mesthrie and Bhatt (2008), Melchers and Shaw (2011).

To conclude, standards in English language have always been dynamic and



constantly evolving over time, “articles continue to be written on this and related topics perhaps in order to ensure that the old ideas do not resurrect” (Bamgbose, 1998, p. 12). WE does not dispute ‘standard English’; what it disputes is monolithic standards like those in the inner circle. Such a multilingual perspective is reflected in the growing acceptance of the key word and concept ‘pluricentricity’, a key feature in the WE paradigm that upholds multiple standards and norms.

Recognition of speaker identity is premised on the acceptance of multiple standards rather than following a monolithic approach. This aspect is especially significant for English learning in the expanding circle, for teachers and learners can choose a more relevant and attainable endonormative, natvised model in ELT practice. As Cook (2002) has pointed out:

If L2 learners feel that the chief measure of L2 success is passing for native, few are going to meet it. Both teachers and students become frustrated by setting themselves what is, in effect, an impossible target. (2002, p. 331)

Thus, it is impractical to set an ‘inner-circle standard’ learning outcome as the goal of CE teaching pedagogy.



2.1.3 Standards and CE

Hence the word ‘standards’ (in plural form) is more desirable in the Chinese context. One key element that constitutes the importance of plurality of standards for CE users is the way they learn English. For the absolute majority of CE learners and users, the main process of and driving force for learning English since the 1980s remain more or less stable, that is, as a school subject for *gāokǎo* (高考)—the qualifying entrance examination for tertiary education, supplemented by the motivation of upward and outward mobility (e.g., going abroad or getting promoted). One notable need in recent years is to use English to absorb information particularly in printed sources, especially from the internet. The nature of such intra-national needs for English determines that CE users are not in great need of achieving ‘inner-circle’ competence. Rather, in the context of CE, the goal post of ‘standards of English’ should include the following competency areas:

- a) maintaining speaker identity: communicating with others effectively (whether L1 or L2 users of English) – without compromising their L1 identity;
- b) ensuring intelligibility in ELF communication: to convey speaker meanings that are maximally free from intelligibility problems;



- c) developing a sense of ownership and pride in the local(ised) variety: to make learners proud of and confident in the local(ised) variety;
- d) being equipped with English for academic purposes (EAP) to facilitate life-long learning: to be literate in, and conversant with, lexico-grammatical features of the written standard variety in order to absorb all kinds of information in print or on the internet. (D. C. S. Li, 2007, p. 14)

Furthermore, the performance indicators of the above-mentioned features of WE to users of English in China should be multiple, including the following two core topics: awareness of linguistic variability and linguistic prejudice.

2.1.3.1 Awareness of linguistic variability

As discussed above, pluricentricity means the dual acceptance of new varieties and their indigenized standards; the first reality learners have to face is that they have to be aware of linguistic variability.

First, users of CE should know that variability is a fact of life when it comes to real language use. One obvious example is pronunciation, as pointed out by Nelson (2011):



It is clear from casual observation and from what we all learned in our required courses in introductory linguistics or basic phonetics that it is at least highly unlikely that *any* two speakers pronounce the same word the same way, or even that the *same* speaker ever pronounce the same word the same way again. This is naturally true of ‘native speakers’ as well as ‘non-native speakers’. That being so, why is anyone surprised that speakers of different varieties of English do not pronounce words in the same way? This very noticing of mismatches between what we hear and what we think we produce ourselves tells us that we do not expect or need *identity* of segmental pronunciations from all participants to make their speech intelligible or to make ourselves intelligible to others. (Nelson, 2011, p.29, italics in original)

One benefit of being aware of linguistic variability is that learners/users can have a guilt-free position without having to worry about whether their English is non-standard, partly because of their limited command of linguistic resources, and partly because they would feel more comfortable when engaged in naturally occurring meaning-making activities.



2.1.3.2 Linguistic prejudice

While an awareness of variability in English helps strengthen one's confidence and sense of language identity, linguistic prejudice is bound to arise, and is something to avoid through awareness-raising.

Getting rid of linguistic prejudice is both an important task for people outside China and inside China. For example, the majority of non-native teachers of English should be aware that when their students are neither fulfilling a “patriotic enterprise” (X. Gao, 2012, p. 351) nor being perfect in English, something more than a standard/norm should be considered in the process of teaching and learning: the lingua-cultural identity of both the teacher and students should also be taken into consideration.

In sum, the WE perspective towards varieties of Englishes seems more suitable as a premise for researching CE features. It provides for the possibility of recognising the standards which suit speakers of English in China, and more importantly, allows CE users a ‘speaker identity’ while learning and using the language at the same time. Following this perspective, the next section will discuss the relevance of the WE perspective to CE in the Chinese contexts.



2.2 Definition of China English

Up to now, several terms have been used by scholars to refer to the English spoken or written by Chinese people. However, there is currently no universally agreed term resulting from this ongoing debate and discussion. Table 2.2 presents some common terms that are adduced from the literature on the functions and status of English in Mainland China.

Table 2.2: Proposed terms in reference to the English used by Mainland Chinese

Term	Source
Chinglish	Niu & Wolff (2003, pp. 9-10); Zhuang (2000, p. 7)
Chinese English	Wang & Ma (2002, p. 56); Kirkpatrick & Xu (2002); Xu (2010)
Sinicised English	Cheng (1992, p. 162)
China English	Ge (1980, p. 2); Jiang & Du (2003, p. 27); He & Li (2009, p.71)
Chinese-coloured English	Huang (1988, p. 47)
Chinese-Style English	Gui (1988, pp.13-14)

Many definitions have been attempted, and a number of writers have argued in favour of one of the above terms or the other. For example, Sun (1989), W. Z. Li (1993), Xie (1995), Jia (1997), Poon (2006), D. Y. He (2011) offer their views toward the most appropriate term and definition, each from their own



perspective. The main arguments in favour of one or the other term will be discussed at length below. One general tendency is that those terms that suggest inferiority and social stigma are less frequently used today, while two terms remain mainstream: China English and Chinese English.

The rendition of either of these two terms into Chinese is not a problem, because both are translated as *zhōngguó yīngyǔ* (中國英語, literally ‘China English’). In line with He and Li (2009), we believe ‘China English’ (with a noun as pre-modifier) is more suitable compared with ‘Chinese English’ (with an adjective as pre-modifier). Thus the term ‘China English’ will be used in the rest of this thesis. However, this term has been used by different Chinese scholars with slightly different meanings (Jia & Xiang, 1997, p. 11; Jin, 2002, p. 72; W. Z. Li, 1993, p. 19; Wang, 1991, p. 3; Xie, 1995, pp. 7–11; He & Li, 2009, pp. 70-89). There is thus a need for clarification.

For the purpose of this study, the term ‘China English’ used here is characterised by two closely related features. In particular, China English is:

- a) “A continuum covering learners’ English and well-educated users’ English”(Gui, 1988, pp. 13-14); and
- b) “A performance variety of English which has the standard Englishes as

its core but is colored with characteristic features of Chinese phonology, lexis, syntax and discourse pragmatics, and which is particularly suited for expressing content ideas specific to Chinese culture through such means as transliteration and loan translation.” (He & Li, 2009, p. 83)

The first feature (a) emphasizes the fact that CE is not confined to any specific group of people because there is no clear cut-off point between a learner and a user of English. In terms of how and how much English is used in their lifeworld (Husserl, 1970), the millions of Chinese people who are learning or have completed their learning of English in formal education may be positioned somewhere on a continuum. The second feature(b) emphasizes the fact that, being a new variety of English, “it is only natural that China English is characterised with cross-linguistic influences from the Chinese language since the learners’ acquisition of a second language is influenced, either negatively or positively, by their mother tongue, and by the linguistic environment” (He & Li, 2009, p. 83).

2.3 The China English debates

While CE is recognised as one important component in the expanding circle of the WE family, and the status of English in China has been a topic of



increasingly intense debate among scholars in Mainland China over the last three decades, there is no generally agreed-upon account of, let alone a research agenda for, the codification of its phonology, lexis, grammar and discourse features, since different researchers define the status of English in China in their own ways. As this is central to the discussion of CE as a developing variety of English, a lot of research has been carried out. In the literature, there is much controversy surrounding the features of China's English. The key questions include the following:

- a) What should we call English in China?
- b) Is there a Chinese variety of English? If not, what stage is it at currently?
- c) Does it have its own features: home-grown lexico-grammatical innovations rather than errors?
- d) Are 'overuse' and 'underuse' of English words in learner English undesirable, and should they be adjusted?
- e) Which English variety should serve as the norm for English teaching in China, and what should its constituent grammar and lexis look like?

The fact that these questions remain unanswered demonstrates that the exact status of English in China is still very much an area of ongoing debate and research, in China and beyond.



2.3.1 Debates in Mainland China

In Mainland China, the debate about China English was initiated by Ge (1980), when discussing the translation of culturally distinctive lexis such as *Four Books*, *eight-legged essay*. Ge argued that these translations are examples of China English, not Chinese English or Chinglish, since the latter two connoted inferiority and poor education at that time. Various alternative terms have been used by different researchers to refer to the same phenomenon: Chinese English, Chinglish, and Sinicised English. Huang (1988) put forward a positive argument towards having a Chinese-flavoured translation. Sun (1989) took these arguments further, analysing English in China from the perspective of a new language variety, although in his view, Chinese English was not yet a variety. Such a debate constitutes evidence of Chinese scholars' growing awareness of the need for fine-grained research on linguistic characteristics specific to English in China, beginning with words that capture the socio-cultural experiences of its English users in China. After the 1980's, there came a phase of more polemical debate, with more specific questions being raised, the first of which was "What should we call English in China, and what is the definition of it?" which is of course open to discussion. As mentioned in Section 2.2, several terms have been proposed in the literature on the topic. Most of the discussions and definitions focus on Chinese-English translation,



until Wang (1994), who defined CE as “the English used by the Chinese people in China, being based on standard English and having Chinese characteristics” (p. 7). The reason why his definition is seen as a milestone in studies of CE is that by this definition, CE is no longer an abstract entity, but a well-demarcated language variety, thus distinguishing its differences from mistakes and this point distinguishes Wang’s contribution from others. Later, W. Z. Li (1993) provided an influential point of view by saying that there are two kinds of English in China: China English and Chinglish:

China English, with its indigenous lingua-cultural expressions, nativised lexicology as well as unique syntax and discourse structures as its major features, contributes much to international communication...it is considered as an accepted variety, contributing much to enriching the English language. Chinglish refers to the Sinicised English usually found in pronunciation, lexis and syntax, due to the linguistic transfer or ‘arbitrary translation’ by the Chinese English learners, thus being regarded as an unacceptable form of English...it is not commonly used, causes barriers and misunderstandings in communication. (p. 80, original translation by Li)

Although W. Z. Li admitted that the two were inter-related, he offered useful



clarification, by characterizing China English as ‘of great significance’ and Chinglish as ‘pathological’ as well as undesirable. In the same article, he revised the definition of CE as:

...the lexis, sentence structures and discourse that have Chinese characteristics. It takes normative English as the core, and it expresses things that are uniquely Chinese. It bears no mother tongue interference, and it manifests in communication by means of transliterations, loan translations and semantic shifts. (p. 19, original translation by W. Z. Li)

Though many others have since joined in the CE definition debate, such as Zhang (1995), Xie (1995) and Jia and Xiang (1997), the question remains an open one. The fact is, for the past three decades, the terms China English and Chinese English have both been used by researchers in the discussion. The most important change in this round of discussions, however, lies in the fact that people began to use the term ‘language variety’ when talking about Chinese English, for example Jia and Xiang (1997), Du (1998a, 1998b), Luo (1998) and Lin (1998). It should be noted that not everybody agrees that it is a variety; rather, scholars simply noted that it is a distinctive, objectively existing form of English being used, and began to realise that it is characterised by its own features, which have yet to be specified and described explicitly in research.



The decade after the 1990's witnessed deep reflections on the previous arguments, but also some confusion. Du and Jiang (2001) presented an overview of 20 years of China English studies, addressing CE as the “newcomer in the family of World Englishes” (2001, p. 79). But this view was challenged by Qiu and Ning (2002), who responded to Du and Jiang’s article and concluded that it is impractical and unnecessary to nativise English in China. They believed that putting forward the concept of CE was inappropriate, since it was likely to interfere with the popularisation of Standard English, and argued that CE phonology, lexis and discourse were unintelligible to those outside China. They claimed that “the existence of China English will contribute to the prevalence of Pidgin English and ossification of interlanguage” (p. 23). Du and Jiang (2002) then responded to the Qiu and Ning (2002) article, and after clarifying some misunderstanding of issues concerning the spread of English as an international language, and the exact meanings of a few applied linguistic terms, they moved on to a review of a number of studies of English as an Asian language, through which they believed “a better understanding of China English is expected to emerge” (p. 27). These three articles generated a lot of heated discussion about attitudes to the English used in China. However, despite disagreements, both sides recognised the importance of investigating CE. Their differences lay in the purpose of the investigation: to eliminate CE or to legitimate it.



Other writers, for example Zeng and Lu (2006), believed we should not be too rushed in our approach to CE, because “only when endowed with sufficient, stable features of pronunciation, vocabulary and syntax compared with the ‘norm’, does a variety take shape. At the moment, the system of China English only exists in fragments, not in a system” (Zeng & Lu, 2006, p. 55, ‘English abstract’). In an attempt to sharpen the focus of the debate, Han (2007) searched 8,209 journals in Mainland China, and found that between 1979 and 2006, there were only 337 articles on the topic of China English. She then concluded that it was a research question at an early stage of its development. Whether CE is a variety or not is still not clear. X. L. Hu (2008a) proposed a functional perspective on the question of a Chinese variety of English by using Kachru’s (1992) three-concentric-circle model and Kirkpatrick’s (2007) classification of language functions. She claimed that CE serves many functions in China. She also stated that one problem in existing research on CE as a variety is that empirical studies of the features are rare, and (up to that point) no corpus of CE existed (X. L. Hu, 2008b). In recent years, scholars in Mainland China with an interest on this topic seem to have reached an agreement on this point: rather than abstract arguments, empirically grounded evidence or proof is needed.

In short, arguments and counter-arguments have accrued during the past two decades; new ideas were embraced by some yet rejected by others. However, it



may be noted that only when the discourse surrounding WE and CE remains controversial and pluralistic can progress be expected. A welcome note of optimism was sounded by X. Q. Hu, who stated that “it is to be hoped that in the course of time China English will become an honored member of the inner circle” (2004, p. 32). Though X. Q. Hu’s optimism may not be transformed into reality in the near future, it helps advance the cause of CE and its further development. One argument for this is the Chinese borrowings in English. J. Yang (2009) reported on a corpus-based study comparing Chinese borrowings with acceptance in general English that appear in ‘desk’ dictionaries published in the 1980s with those listed in dictionaries used at the time of writing, and found that lexical items borrowed from Chinese have not only increased by about 54 per cent (from 100 to 154 in numbers) during the past 20 years but also occur more frequently. When put into semantic fields, these added new words could be classified into the following categories: philosophy, religion and literature (e.g., *feng shui*, *Li Po*); history and politics (e.g., *laogai*, *Shang*); food and drink (e.g., *tofu*, *hoisin sauce*); ethnography (e.g., *Miao*, *Tai*); martial arts (e.g., *qi gong*, *wushu*); and other (e.g., *pipa*, *Putonghua*) (J. Yang, p. 99). His study also spotted a trend that the newly added words are mostly from Mandarin (i.e. Pinyin-based), rather than Cantonese or Amoy, which are responsible for a larger portion of the words found in Cannon (1988). Yang’s study predicted that the borrowing of Chinese words will be expedited, and that there will be more



Pinyin-based Mandarin loanwords falling into various semantic fields. This is indicative of a trend whereby Chinese borrowings into English will continue to expand and enrich the lexicon of CE.

2.3.2 Debates beyond Mainland China

The status of CE as a new variety has also been a subject of academic inquiry outside Mainland China. One much-discussed example was the special issue of *World Englishes* in 2002 entitled *English in China: an interdisciplinary perspective*, where 13 papers present a range of views on a number of important issues around CE. Bolton's paper, which examines the history of English in China from the seventeenth century to the present, is one of the most influential and thorough accounts in the anthology (cf. Bolton, 2003). Pang, Zhou and Fu's paper (2002) reports on a survey analysing the impact of China's entry into the WTO on the linguistic behaviour of people in China. A total of 400 people in 126 companies in five cities were interviewed, with questionnaires containing issues related to the learning and teaching of English. Results show that China's entry into the WTO could be a powerful force for change as far as people's attitudes to English language use and teaching are concerned. Most respondents to the survey believed that English is becoming more important in the context of the WTO, that speaking and listening should be the focus in the future, and



that people's motivation for learning English would become more used-oriented, in addition to being a school subject. This is the paper's most crucial finding, since it demonstrates the urgency felt towards a more pragmatic approach to English language teaching in China. However, other interesting results also emerge from this report. One of them is that most people interviewed expressed their concern about the ownership of English, saying that English should not be owned by British or American people. Rather, users of English should share equal rights in this communication tool. Adamson, in both *Barbarian as a foreign language: English in China's schools* (2002, pp. 231-243), and in his later work *China's English: A history of English in Chinese Education* (2004), presents another thorough historical account of English in China's schools, from the mid-nineteenth century to the present. He describes and analyses systematically a number of factors related to English teaching, for example the official role and status of English for the government and the people in the domain of education at different times. The issues included for scrutiny range from the curriculum to the contents of textbooks, making this the most useful account of the development of English language teaching in China over the past 150 years.

Adamson's works are widely held to be authoritative resources for CE research (Kirkpatrick, 2007; Xu, 2010) for two main reasons. Firstly, what Adamson



accomplished is a systematic, detailed and chronological study of the factors influencing English language teaching in China, based on his personal involvement as an ELT consultant in the field, as well as on instructive documents that he has brought to light. Secondly, compared with some other reports of this kind written by Chinese people from Education Bureaus at different levels, his account is more personal and subjective, since he did not need to worry about undue administrative consequences for criticizing the government policy. He found clear trends suggesting that the current attitude towards English is the most positive one in Chinese history. This conclusion is echoed by Lam's (2002) interview with 407 people from Mainland China. Responses show that learners of English are now experiencing more favourable learning circumstances, with a steadily improving climate for English teaching and learning. In the same issue of *World Englishes*, a bibliography of English in China (Adamson, Bolton, Lam & Tong, 2002) is included for reference. One final concern centred on CE relates to people's views of it, particularly their "concern for lingua-cultural identity" (D. C. S. Li, 2009, p. 82). Kirkpatrick (2006) comments on this problem, arguing that it is "curious that the views and voices of millions of real 'consumers' of ELT are seldom consulted and represented in research" (p. 72).



2.4 Learning and teaching English in China

Eight years ago, in the analysis of a report written for the British Council called *English next*, Graddol (2006) predicted that “the growth of China would have a significant impact on the world in which English was used and learned”(Graddol, 2006, p. 9). The following section will review a body of relevant literature with a view to teasing out other related factors behind English learning and teaching in China.

2.4.1 Thirty years of the CET4 test

The 1980s was important not only for witnessing the debate discussed above, but this was also a critical period for research on CE features, since the language users, the people who could provide evidence for or against in the debate over China English, emerged in significant quantities only after the mid 1980's. The reason for this was that prior to the Cultural Revolution (1966-1976), language teaching in China focused on Russian, not English, and during the decade-long Cultural Revolution, the education system (English language teaching included) was paralysed by political forces. In 1979, the Ministry of Education decided that foreign language teaching, especially in English, needed reinforcing. The following recommendations were made:



- a) English teaching in primary and middle schools should be intensified.

In middle schools, English language should be the basic course, alongside Chinese language and mathematics. In primary school, from third grade onwards, all pupils should learn English.

- b) English teaching in universities should be strengthened. All university students should take a common foreign language course. Every student in all subjects (including the Human and Natural Sciences) should learn one foreign language (English, Russian or Japanese), while the foreign language universities should become the key providers of foreign language education.

- c) Training of teachers of foreign languages should be intensified. Teachers should be sent abroad to learn foreign languages, and foreign experts in linguistics should be invited to China to give lectures.

- d) Research into approaches to language teaching and applied linguistics should be intensified. (Feng, 2007, p. 5)

A new curriculum for English language teaching was not issued by the government until 1980 – almost 20 years after the last one in 1962 – which marked the beginning of the era when English became highly desirable for communication, and was strongly promoted in schools by the government. *The College English Curriculum* (1986) was published as a guideline for university



students in China, and was based on the first corpus in China, *JiaoDa English for Science and Technology* (JDEST) (Yang, 1983). It was a milestone in that it clearly stated, for the first time, that the purpose of learning and teaching English is spoken and written communication, and the vocabulary list based on that corpus was explicitly attached as the teaching target. It began to talk about setting up criteria for and expected learning outcomes of English proficiency for university graduates (the College English Test, Band Four, hereafter CET4) and stated that to pass this test, students should have a vocabulary size of 3800-4000 words. The first CET4 test was held on 20 September, 1987. Of the 102,821 people who took the test, 53,871 passed (Li, Zhang, & Liu, 1988, pp. 560-561). From that point on, CET4 test became a compulsory test for university graduates regardless of their major discipline and remains so today.

One point arising from these details is particularly salient here: the people who took the CET4 test in 1987 are now in their 50's, and those who took the 2012 CET4 test are in their 20's. These people aged from 50 to 20 account for the majority of potential CE users, both now and into the future. One reason why they may be viewed as such is their age; the other would be that they are likely to have more chances to communicate and contact with other users of English in China and beyond, which is a natural part of their lives in the 21st century following the popularisation of increasingly easy access to the internet, mobile



phones and other electronic devices. One may argue that people do not learn English solely from schools. However, school education would certainly constitute the main source of language input for the majority of Chinese: the set of textbooks from primary school to universities, teachers, homework, and countless test papers. For most learners of English in China before the internet era, these represented the only resources that gave them exposure to English, a window through which they got to know the outside world.

Only around 0.1 million students took part in the first CET4 exam in 1987. By 1998 the number had grown to 2 million each year (Yang, 1998) and reached 4.48 million in the year 2002. The total number of attendants during the period 1978-2002 added up to around 23 million people (Yang, 2003). These numbers keep growing as universities in China are admitting more and more students. It is no exaggeration at all if we estimate that the past three decades produced more than 100 million graduates who passed the CET4 exit benchmark. In 2004, the number of test-takers reached 11,000,000. It is estimated to be “the largest scale English test in the world” (Feng, 2007, p. 5). After passing the CET4 test, some students go on to take ‘voluntary’ tests in English such as GRE, TOEFL, IELTS, etc., for various reasons ranging from studying abroad to emigration. Though not all of them pass it, the exam-takers go through essentially the same process of learning English on a compulsory basis in school settings, and



develop their own variety of English in comparable ways, in similar language environments and in response to similar stimuli. It is the learning process that enables them to become English users, not the examination.

A common argument arises in response to this: their English is not very good, and the term ‘knowing English’ has plenty of latitude. This is a linguistically prejudiced statement using a norm (British English or American English usually) from the inner circle country, one which may not make much sense nowadays. British English and American English are only varieties of WE, not necessarily the norm and standard that the world must look up to. “The idea that there is some form of fixed standard of a language that everyone who speaks the language always uses in exactly the same way leads people to misunderstand how language works in real life” (Kirkpatrick, 2007, p. 12).

University graduates are Chinese users of English, and they are increasingly numerous in number. And this trend will continue, as university applications by Chinese students are increasing by 8% each year (Cai, 2004, p. 19). About 1.08 million students were admitted to university throughout China in 1998, and in 2002 the number went up to 3.2 million. The total of in-school university students in 2003 alone was 17 million, and by 2020 the estimated nation-wide target is 40 million (Cai, 2004, p. 19). If these numbers of students are extended



to include all students of English in China, it would have outnumbered the population of the United States. Estimates of the number of English learners and users in China varied between over 200 million and 350 million (Bolton, 2002; Kirkpatrick & Xu, 2002; Xu, 2010).

Of course, mere numbers mean little if we insist on using ‘standard English’ as a reference. However, considering the impurity of British English (Kirkpatrick, 2007, pp. 39-54) itself, ‘China English should stand alongside British, American, and the other ‘world Englishes’ (X. Q. Hu, 2004, p.26) as a norm. Evidence-based proof need to be established before CE can be regarded as an autonomous, legitimate variety in its own right. To put the number of CE users in a broader context, we need to situate it in a larger picture. For instance, consider the estimated numbers of speakers of English and Mandarin in Tables 2.3 and 2.4.

Table 2.3: Speakers of English at the start of the 21st century

Native speakers	Other speakers
(English as L1)	(English as L2 or additional language)
375 million	1,125 million

(Graddol, 1997, p. 10)

Table 2.4: Distribution of proficient speakers of English and Mandarin worldwide (2002)

Rank	Native speakers	Million	Rank	Competent speakers	Million
1	Mandarin Chinese	800	1	English	1900
2	English	350	2	Mandarin	1000

(Time Magazine, 24 July 2002)

These figures show that more and more people other than native speakers are using English and they have outnumbered the traditional native speakers. They also reveal that Chinese and English are two linguistic giants, and that millions of Mandarin speakers are bilingual speakers of English (i.e., CE users). These statistics suggest that to ignore the features of CE would be to neglect the mainstream variety of the future. With several hundred millions of competent speakers of English constituting an important part of ELF communication in the world, this amazing fact is obviously going to have a dramatic effect on the forms this chameleon language takes on. For example, in the field of scientific research, there is a strong need for academia to better understand how CE features, the way English is used by Chinese bilingual users, in order to promote mutual understanding as well as avoid cultural conflicts that are likely to be occasioned by different interpretations of the same topics being discussed.

About the number of CE users, precise figures are lacking, but Bolton and



Graddol (2012, p. 3) estimated that there are up to 400 million learners overall, about one third of the entire population, and some 50 million schoolchildren are learning English right now, with figures rising continuously. Wen (2012, p. 81) cited statistics from 2008 that counted 163 million students in formal instruction, while Xu (2010, p. 204) assumes a “guesstimated 10 to 70 million competent Chinese users of English”. Wei and Su (2012, p. 11) report hard statistics from a national language survey conducted around the turn of the millennium, yielding a figure of 390.16 million learners of English. The figures in Table 2.5 constitute another example of the size of the number of Chinese people who are learning English.

Table 2.5: The English learner populations in six major cities in China (2007)

City	Approximate total number of learners
Beijing	3,814,200
Shanghai	2,047,800
Guangzhou	1,502,100
Chengdu	1,110,300
Wuhan	902,700
Shenyang	822,300
Total	10,199,400

(Source: *China Statistical Yearbook*, 2007)

Besides finishing the requirements of English as a school subject, the huge



number of people who learn English may have other plans, too. One of these may be studying, working or living abroad, with America and U.K being the two most-wanted countries or popular destinations. Figure 2.3 gives some indication of how many of them are doing so.

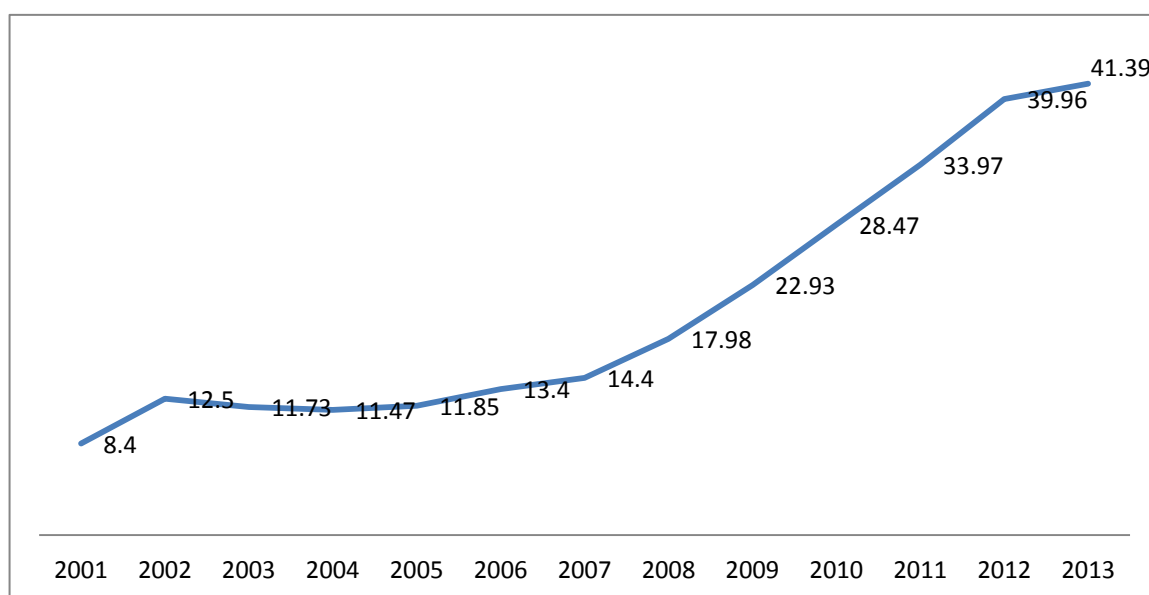


Figure 2.3: Number of Chinese people studying abroad (2001-2013)

(in tens of thousands)

(Source: *Annual report on the development of China's study abroad*, 2013)

Studying abroad has become increasingly popular among Chinese students. Even the high tuition and expenditure standards cannot stop their enthusiasm of studying abroad. The destinations of these people varied, but among them the United States and the United Kingdom are the top two countries receiving Chinese students. Figure 2.4 shows the statistics of students studying in United States from 2007 to 2012.

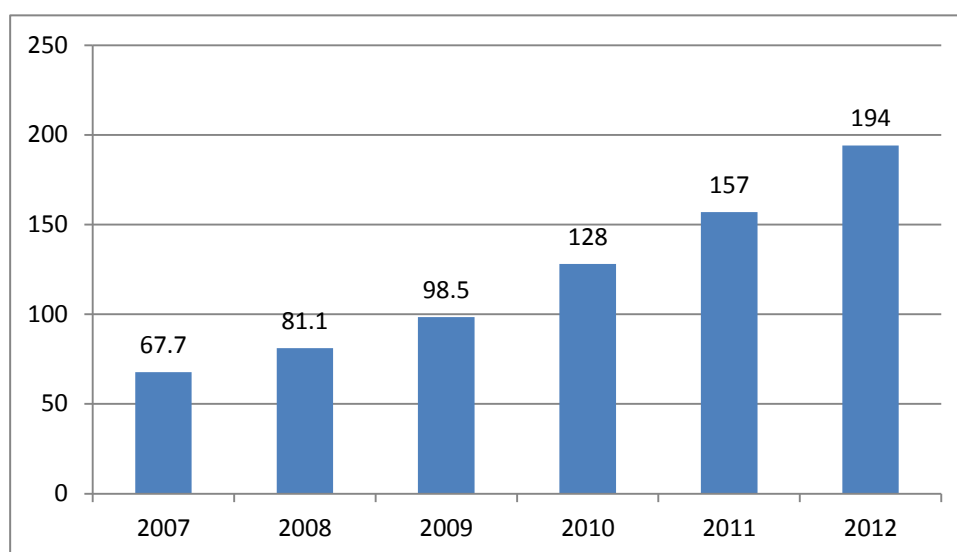


Figure 2.4: The statistics of Mainland Chinese students studying in United States (2007-2012) (in thousands)

(Source: *Annual report on the development of China's study abroad*, 2013)

The statistics of students studying in United Kingdom from 2007 to 2012:

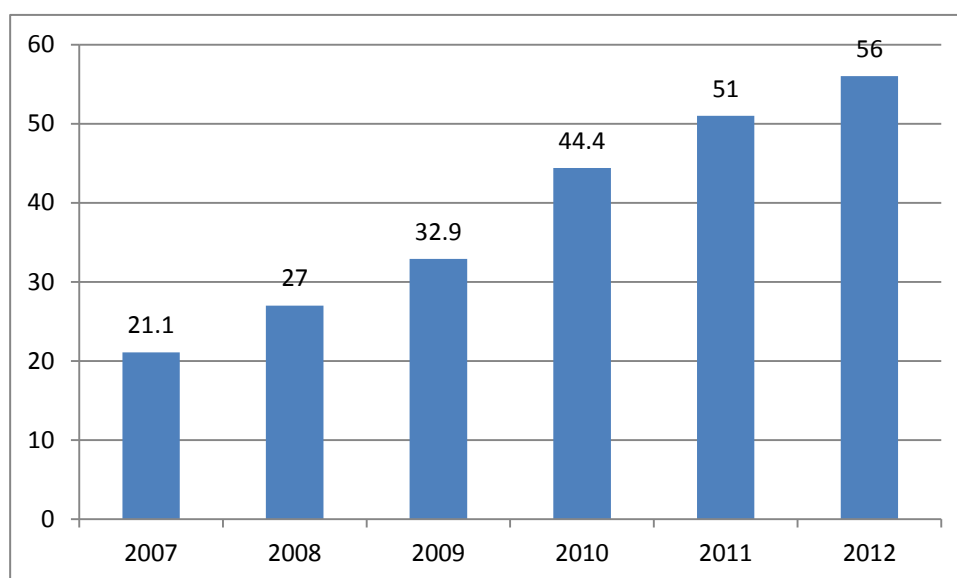


Figure 2.5: The statistics of Mainland Chinese students studying in United Kingdom (2007-2012) (in thousands)

(Source: *Annual report on the development of China's study abroad*, 2013)

These trends suggest that changes have occurred to people who are learning English in China, especially in the new century characterised by globalised economies and education. In short, English learning in China may be associated with or driven by one or more of the following demands:

- a) the development of basic skills (e.g., how to use the computer, including spreadsheets, and conduct searches on the internet) to meet the generic demands in information communication technologies (ICT) needed to acquire new knowledge and specialist skills in the future, namely learning how to learn;
- b) an indispensable part of fundamental preparation for further education abroad; and
- c) the mastery of various ICT tools for international communication.

In a word, CE users may be characterised as practitioners of “English-knowing bilingualism” (Jenkins, 2007, p. 163; cf. Kachru, 1992) whose use of English deserves to be better understood through empirical, systematic research.

2.4.2 The non-native teachers

English teachers in China constitute one key factor that has significant influence



on the language learning outcome for millions of learners in China. In other words, if the English-knowing bilinguals are products, the non-native teachers of English are the moulds. Their teaching and guidance, along with their language use (some combination of Chinese and English in reality) will inevitably exert considerable influence on learners at the formative stage of their own variety of English.

According to the *Law for Teachers* (*jiàoshī fǎ* 教師法), the term ‘teacher in China’ is first defined and limited to those whose nationality is Chinese. Theoretically no foreigners can be teachers of English in state-owned schools in Mainland China. So virtually all teachers of English in state-own schools in China are non-native speakers of English. Salient terms of employment of the two types of teachers are outlined in Table 2.6.

Table 2.6: Two types of teachers in China

Teachers	Administrative branch	Terms of employment
Chinese	Education Bureau of PRC	official teacher (gōnglì jiàoshī 公立教師), life-long contract
Foreign	Foreign Affairs Bureau of PRC	‘foreign expert’, short-term conditional contract

Nowadays, though more and more ‘foreign experts’ come to China to teach English, their contracts are usually short (1-2 years). I have been a learner of English for 25 years and a teacher of English for over 15 years in Guangdong Province, Mainland China. My observation of foreign teachers of English is that they usually teach Oral English or Writing in schools, which typically consists of one class per week. The goal of their teaching is to complement the ‘normal teaching’ (*chángguī jiàoxué* 常規教學) of English which will be one of the major subjects in all school levels from primary to university. ‘Normal teaching’ is organised and scheduled nationally by the Education Bureau, with a strictly regulated curriculum and textbooks, followed by national English proficiency tests. The foreign teachers, on the other hand, do not follow a school curriculum since they are not part of the system. Most of the time, therefore, they can choose their own teaching targets, using their own preferred teaching methodology. The tests they set for students do not affect the students’ overall performance. One salient phenomenon these days for foreign experts teaching English in China is that they teach in the private sector, in language schools which do not belong to the public school system but to the private language teaching and learning market. However, compared to the millions of learners in state-owned schools receiving compulsory education in China, the number of customers who choose to add to or consolidate their English proficiency in private language schools for whatever purpose, is relatively small.



For example, according to the *China Statistics Yearbook, 2010*, in Guangdong province alone, there are around 30 million school-based students. Thus, for the absolute majority of learners in China, so-called non-native teachers constitute the backbone of English language teaching and learning. The non-native teacher is a central feature of and key player in the English learning process in China. There is no official data about the total number of English teachers in China, although according to Bolton (2003), the number of English teachers has increased from an estimated 850 in 1957 to well over half a million currently. However, this is certainly an underestimate, since Dai (2002, p. 11) reported that at the time of research there were over 474,000 teachers of English in secondary schools in China, and in 2009 he wrote that there were over 800,000 in-school English major students aiming to become English teachers after graduation (Dai, 2009, p. 2). English majors in universities are normally expected to become non-native English teachers in China, while foreign teachers from English-speaking countries would usually teach oral English, but not the mainstream classes which account for a greater part of students' time and exposure to English in school. It is likely that the way the non-native English teachers go on to teach will reflect typical features of their own learning. With such a process occurring over a long period of time impacting millions of learners and users of English, a new variety of English may well begin to emerge. It is likely, therefore, that as a consequence of the growing popularity



and use of English in China, both intra-nationally as a developing variety of the *WE* family, and inter-nationally as a *lingua franca*, it will probably evolve at a faster pace than other varieties of English. If we consider English and Chinese as two elements that may evolve mutually when in contact with each other, there is good reason to believe that the internet functions as a catalytic agent. According to *The 33th Statistical Report on Internet Development in China* (issued 2014 January), produced by the *China Internet Network Information Centre* (CINIC, www.cnnic.net.cn), the number of netizens in China reached 618 million by the end of 2013, with an increase of 233 million since the end of 2009. Among these people, around 500 million are mobile phone internet users. This might well be one reason accounting for Tom McArthur's statement:

Not everyone interested in English needs to give much thought to Chinese. This statement is so true that it borders on the banal, and a similarly true and banal statement might be: Many people involved with Chinese don't need to give much thought to English. However, things are changing, and fast. (McArthur, 2003, p. 2)

English for Mainland Chinese people is more than a school subject, due to the advance of the information technology revolution and a compulsory education system which makes people bilingual.



2.4.3 CE versus interlanguage

Language can be looked at from different perspectives. One example of this is the use of terms such as ENL, ESL and EFL as opposed to the ‘Three Circles’ model, both being models to categorise different varieties of English. Another is interlanguage versus variety of English (for example, CE), and there are a number of interesting discussions of this issue. For example, D. C. S. Li (2000) presented three main arguments in support of his view that English in Hong Kong was an interlanguage. First, English in Hong Kong is rarely used among Hong Kong Chinese for social interaction in informal settings—for intra-ethnic communication, which means that Hong Kong English does not have the linguistic environment in which to develop as a new variety; second, indigenous norms have not developed; third, English in Hong Kong is exonormative since standard English as codified in grammar and dictionaries, and manifested in ELT materials continue to be the norms of reference. Thus Li concluded that “it is more appropriate to characterize Hong Kong English (HKE) as an Inter-language rather than a new variety of English” (2000, p. 57). Nevertheless, communication needs and modes around the world have rapidly changed beyond all recognition in the past decade, and things are not the same anymore. In line with many scholars (Cook, 2002; X. Q. Hu, 2004; Graddol, 2006; Jenkins, 2006; Schneider, 2014), I also believe that it is high time that we



revisited this ‘interlanguage or variety’ issue.

As a starting point, it is worth reviewing the origin of the concept ‘interlanguage’, and the circumstances under which it was proposed. As far back as the 1970s, Selinker (1972) put forward a claim that learners of English produced their own learner language variety, which could be examined as a system with its own rules and characteristics. “The term was coined by Selinker in the belief that the language learner’s language was a sort of hybrid between his L1 and the target language” (Corder, 1981, p. 2). This is a kind of language that second language learners develop. Such studies are usually referred to as ‘interlanguage studies’ or the study of ‘learner language’. This concept was suggested in order to draw attention to the possibility that the learner’s language may be regarded as a distinct language system with its own particular characteristics and rules (Stern, 1983, p. 125). Whether it is right to consider the learner’s language as a ‘language system’ is debatable, but the attempt to do so illustrates “the linguist’s intention of understanding all kinds of language varieties” (Corder, 1981, p. 2). The interlanguage hypothesis uses standard English as the norm or target language. If a learner does not reach that target, s/he is still a learner by definition. Selinker also suggested that on average only about 5% of learners could reach the target language proficiency level (Ellis, 1985, p. 48), which implies that 95% of learners would never get there, because



their linguistic competence would reach a plateau with progress being no longer possible, a stage of learner language development known as fossilisation. This is one of the reasons why the concept of interlanguage is so significant: most of the learners and users of English fall into this category! Selinker and his followers advocated studies of interlanguage as a system in its own right, noting that its structure can be observed by studying the utterances of learners who attempt to make meaning in the target language.

In sum, the notion of interlanguage was an intermediate goal in the learning process. The positive side of proposing this construct is that it provides a theoretical framework for interpreting Second Language Acquisition (SLA) as a cognitive process and allows us to look into language learning processes and outcomes. Moreover, it provides a framework for investigating learner language and the learning process empirically. However, two questions are often asked:

1. What/which one is the target language? From the discussion of the uncertainty of standards in inner circle countries in section 2.1, it might seem that there is no single target language for all the millions of learners in the world, or at least the standard is dynamic in terms of contexts and purposes. For example, should a Japanese learner of English share a targeted proficiency level with that of a Nigerian learner, and should we



suppose that the goal of English learning for them both is to speak alike? There are exceptions, however, for English for Specific Purposes (ESP) learners. It is quite sensible for nurses and doctors to learn one Standard English variety for career development. But what about those who only learn English for getting in touch with the world—for information, for entertainment and for sharing feelings and news with friends all over the world? These people constitute the majority of English speakers in the 21st century. On the other hand, if only 5% of all learners could get to that ‘native-like’ target, as the interlanguage theory suggests, we might begin thinking about recognizing the inter-(*inter* in the word *interlanguage* suggests halfway or in-between) as the target itself. Then we have to look at ‘interlanguage’ from another angle. In an age when “bilingualism is a normal, rather than special, condition” (Graddol, 2006, p. 117), the usual practice, a native-speaker model as target language, would cause problems for non-native teachers too. Insisting on a native-speaker model means that teachers are required to teach a model which they themselves do not speak and this severely undermines their sense of self-confidence (Medgyes, 1994; Cook, 2002, Seargeant, 2012). It also ignores or negates the linguistic and cultural resources that the local bilingual teachers bring to the classroom, not to mention other socio-cultural issues such as linguistic imperialism (Phillipson, 1992; 2009) and prejudice against other varieties of English.



English curricula, therefore, should not be designed to measure learner performance using the yardstick of native-speaker-based standards. Such a model for learning and teaching is “irrelevant and unattainable” (Kirkpatrick, 2007, p. 376); instead, a “deficit and disempowering” model should “be replaced with a model of difference, whereby learners’ L1 identities and ownership of English are both respected” (D. C. S. Li, 2007, p. 11).

2. Are all the features of an interlanguage undesirable for its users? In search of features of an interlanguage, linguists have found a number of lexical, grammatical and discourse features over the years. If we agree that there is a lack of a single standard for all learners, then we might need to think twice before asking learners to shake off their interlanguage features and try their best to approximate the target language, especially after those features are widespread. Gee (1999) suggests that “non-language stuff such as clothing, manner, gestures, tools and technologies, combines with language in use, establishing identities” (p. 6). It is not very common for learners of English to aim to write and sound like Anglo-Americans when they start to learn English. Indra Nooyi, the CEO of Pepsi since 14th August, 2006, was born in India. She graduated from Yale University and has lived in America for a long time. However, she would always wear a *sari*, the traditional Indian



dress for women in India, for important occasions. This is one example of people establishing identity with clothing. Tommy Koh, a senior figure on the Singaporean government (and one time Ambassador to the UN) says “When I speak English I want the world to know that I am a Singaporean” (Kirkpatrick, 2007, p. 144). This might well be a representative view for people who wish to keep their distinctive features of English because it will give them a sense of identity, along with non-linguistic characteristics. Although “The criteria to use in distinguishing between features of learner English or an interlanguage and features of a new variety of English remain elusive” (Kirkpatrick, 2007, p. 144), this discussion suggests that the features of CE, when they are distinctive from other varieties of English in a culturally systematic way, are actually a distinct variety, not an interlanguage, because of the millions of speakers who use it throughout the world, very often with errors, but nonetheless comprehensibly, but also because of the increasingly important role that English plays. It has only taken thirty years for the knowledge of English to expand among the general population of China. At what pace will it evolve in, say, another thirty years? It may well have become a powerful variety. In this connection, it is hoped that the CE data in this research will come to be viewed as “English with its own particular identity and ownership” for Mainland Chinese users (Kachru, 1986, p. 31).



If we consider ‘interlanguage’ from the perspective of second language acquisition, problems exist too. As Cook (2002) has pointed out: “If L2 learners feel that the chief measure of L2 success is passing for native, few are going to meet it. Both teachers and students become frustrated by setting themselves what is, in effect, an impossible target” (p. 331). Lots of research findings have shown that CE has her own features in several ways. For lexical items, *guanxi* (roughly interpersonal relations or connections or backdoor practice) is one such case which has been exemplified and discussed at length (Gu, 2002; cf. Xu, 2010, pp.153-156), showing how the lexical meaning of an ‘English’ word is fully embedded in the lingua-cultural context of Mainland China. For pragmatic competence, D. C. S. Li (2002, pp. 579-585) used four speech acts to illustrate a characteristically Chinese choice of socio-pragmatic norms in intercultural communication, norms which are guided by traditional Chinese, partly Confucian values:

- a) Terms of address: preferred choice amongst many possible ones;
- b) Responding to a compliment: denying compliment to show modesty;
- c) Expressing disagreement: refrained from openly disagreeing with others;
- d) Readiness to apologize: a useful lubricant that helps build up harmony in interpersonal communication.

As for syntactic features of CE, Xu (2010) identified the following based on empirical CE data collected from Chinese newspapers in China and literary works of the prolific writer, Ha Jin:

1. Adjacent default tense ;
2. Null-subject/object utterances;
3. Co-occurrence of connective pairs;
4. Subject pronoun copying;
5. Yes-no response;
6. Topic-comment;
7. Unmarked OSV sentence structure; and
8. Inversion in subordinate finite *wh*-clauses.

In sum, the discussion surrounding the ‘interlanguage’ or ‘variety’ debate in reference to English used beyond the inner circle is a debate between supporters who adhere to the ‘deficit’ as opposed to the ‘difference’ perspective (e.g., HKE, CE, etc.). Variations of English used by people in the expanding circle are perceived as ‘linguistic deficit’ according to the SLA tradition while the same are viewed as ‘linguistic difference’ from the WE perspective. Both, as is shown in the classification models of Englishes above, are solidly grounded and supported, subject to one’s acceptance of the attendant premises. Our purpose



here is not to resolve the debate; rather, it intends to foreground the existing scholarship that focuses on the English as used among Mandarin Chinese users.

It is important to distinguish between two broadly defined types of context: assessed versus non-assessed. Where English output is assessed (e.g., as part of language tasks in class), CE thus produced is typically learner language, therefore deviations from Standard English, especially lexico-grammatical ones, are more appropriately seen as errors.

On the other hand, for non-assessed situations, deviations may not be seen as errors for the same user may be using English in specific ways out of concern for socio-cultural relevance and speaker/writer identity. Since the focus is meaning-making, correction would seem to be inappropriate and unnecessary. One implication is that interlanguage and varieties of English may essentially be the same thing, the main difference being the context where it occurs and the perspective we take when examining essentially the same language produced by the same person.

So if this thesis argues that CE is a kind of variety of English under development in the expanding circle of WE, it does not imply that interlanguage theory is unimportant or uninteresting. It is one of two complementary



perspectives (CE errors in the learning process still need to be corrected). The English used by Mainland Chinese people may well be an interlanguage if looked at in one way, but the same interlanguage feature may well be part of an emerging variety of English when looked at in another way.

The above views are highly controversial, witness the two rounds of heated debate published in the *Modern Language Journal* in 1997 and 2007. Firth and Wagner (1997) challenge the notions of native speaker (NS), non-native speaker (NNS), and interlanguage (IL) in the second language acquisition (SLA) tradition in their well-known article ‘*On discourse, communication and (some) fundamental concepts in SLA research*’:

[the notion of interlanguage]...conceives of the foreign languages speaker as a deficient communicator struggling to overcome an underdeveloped L2 competence, striving to reach the ‘target’ competence of an idealised native speaker. For the most part, they are applied and understood in an oversimplified manner, leading to an analytic mindset that elevates an idealised ‘native’ speaker above a stereotypicalised ‘non-native,’ while viewing the latter as a defective communicator, limited by an underdeveloped communicative competence. (p. 285)



As might be expected, intense, fierce arguments and counter-arguments ensued. The disagreements were far from resolved, so ten years later, in 2007, the *Modern Language Journal* organised another round of debate on these topics. Looking back on these debates which lasted for over ten years, one may find three camps with regard to these highly controversial viewpoints:

- a) Scholars with unreserved support for Firth and Wagner (1997) (e.g., Block, 2007; Canagarajah, 2007; Kramsch & Whiteside, 2007; Lantolf, 2007; Mori, 2007);
- b) Scholars who are totally objected to Firth and Wagner (1997) (e.g., Long, 1997; Gass, 1998; Gass, Lee & Roots, 2007; Gregg, 2003; Kasper, 1997; Poulisse, 1997); and
- c) Scholars who partly supported Firth and Wagner (1997) but who refrained from taking sides pending more convincing empirical evidence in favour of either side (e.g., Ellis, 2000; Larsen-Freeman, 2007; Swain & Deters, 2007; Tarone, 2007).

In their 2007 article, '*Second/Foreign language learning as a social accomplishment: Elaborations on a re-conceptualized SLA*', Firth and Wagner (2007) describe these two rounds of debate as "fruitful and lively", although for some people, "it was an unwarranted, misguided, perplexing, and naïve critique

of a well-defined field of study” (p. 800). Indeed, the year before these debates, similar assumptions about the distinction between NS, NNS and IL were questioned in Brown, Malmkjaer and Williams (1996), and Cook (1996). The following is one of Cook’s provocative statements on this position, which this thesis follows:

The goal of L2 acquisition should be seen as something other than monolingual native competence...There is no assumption that this knowledge corresponds to a monolingual native speaker’s in either L1 or L2...The starting point should be what L2 learners are like in their own right rather than how they fail to reach standards set by people that they are not by definition. (Cook, 1996, p. 64)

The purpose of reviewing the debate in China and beyond is to show that such views about the absolute distinction of NS, NNS and IL were not and are not universally held. Another example, engaging with the related issue of reassessing our view of English varieties and English as a communication tool, can be found in *The Australian Review of Applied Linguistics* in 2008, which took the debate up to the next level. In that issue, Clyne and Sharifian (2008) first put forward a position paper addressing fundamental challenges and possibilities that require attention, especially the issue of norms. They argue



that the main obstacle to acceptance of other norms than those of the inner circle is attitudinal, and that “a more symmetrical understanding of the pluricentricity of English...requires a mindset appreciative rather than fearful of diversity” (Clyne & Sharifian, 2008, p. 12). The above critical discussion and analysis of interlanguage and pluricentricity may be summed up in Table 2.7.

Table 2.7: Contrasting SLA with WE vantage points towards CE

	SLA	WE
Perspective on deviations	Deficit	difference
Metaphor	Transfer/interference	Language contact/evolution
Norm	Monocentricity	Pluricentricity

In short, the literature surrounding the interlanguage (SLA) versus new variety (WE) debate may be regarded as attempts to unsettle the theoretical premises behind each other’s views, taking into account the complexity of the global spread of English itself, plus the changing contexts where it is being used. The thesis departs from the latter viewpoint, the WE perspective, yet adding some key criteria to the scope that need to be taken into consideration as an integral part of the CE model.

In particular, the forms occur in the speech/writing of CE speakers/writers who have attained a certain level of English proficiency (e.g., through education as assessed in their performance in public examinations), and their CE outputs typically share two characteristics: a) patterned and systematic lexico-grammatical features, and b) being correctly understood without causing any communication problems among fellow CE users.

2.4.4 The status of English in China

Although the nature of Standard English – as a social variety, may be quite stable over time, its role and status in China have gone through dramatic change since the foundation of the People’s Republic of China in 1949. As discussed above, English is mainly taught and disseminated through the formal education system in China. It is used for more purposes than ever before. Vocabularies, grammatical forms, and ways of speaking and writing have been influenced by technological and scientific developments, economics and management, literature and entertainment, in and outside China.

Universities using English as the medium of instruction (MoI) are gradually expanding. For example, at Sun Yat-sen University, a comprehensive university in Guangzhou, the MoI of 298 courses from 2004 to 2010 were officially



characterised as ‘bilingual’ (‘雙語教學’), which means that textbooks are in English and the language of teaching and learning is English, supplemented by some Chinese where appropriate (source: www.sysu.edu.cn/2012/en/index.htm), thereby creating a generation of English-knowing professionals. Economic developments since the 1980’s have led to the rise of a middle class, a group that is more likely to use at least some English on the job. The status of English in China has undergone some perceptible change over time. Adamson (2002; 2004; 2014) has provided useful accounts and insights of the development of English language teaching in China which reflect changing attitudes towards English (see the chronological overview adapted from Adamson, 2002 in Table 2.8).

Table 2.8: Roles and status of English education in China (Adamson, 2002, p. 232)

Period	Role and status	English education
1949—60	Political events made English less favoured; used only as a vehicle for gaining access to Western science and technology; low official status	In the curricula of very few secondary and tertiary institutions
1961—66	The ‘First Renaissance’: popularity of English increased as political pressure waned; English perceived as valuable for modernisation and building international understanding; medium/high official status	Promoted in the curricula of secondary and tertiary institutions

1966—76	Cultural Revolution: English speakers branded as traitors and punished; Western cultural artifacts attacked; low official status	Removed from the curricula of secondary and tertiary institutions; restored sporadically
1976—82	Tentative recovery from Cultural Revolution; English seen as useful for national modernisation; medium/ high official status	On the curricula of secondary and tertiary institutions
1982 onwards	English seen as highly desirable for national modernisation; important for social, academic and economic success; opportunities for study and travel abroad; especially promoted in key schools; high official status	Strongly promoted in the curricula of secondary and tertiary institutions

The timeline shows that the role played by English in China is closely related to that part of history of China herself. English learning and teaching are evidently influenced by these official ideologies. For example, students in the 1966-76 period did not learn much except slogans such as ‘long live Chairman Mao’; ‘good good study, day day up’ and greeted each other as ‘Comrade Li,’ ‘Comrade Zhou’. However, these slogans are not comprehensible to young people in China nowadays, because the English they learn today is totally different, and the role played by English is not the same.

All these changes reflect China’s domestic sociopolitical climate and international relations over time. In fact, “The current attitude towards English



is the most positive in Chinese history” (Kirkpatrick, 2007, p. 145). Further to this, the millions of students who are taking English as a compulsory course from Primary 3 onwards demonstrate the importance of English from another perspective. Apart from being a subject in the school system, English language teaching in language training centres is a new industry in China: the *Wall Street Institute* is a famous English Teaching organisation, which now has 14 branches in 3 Chinese cities. As of 2006, the number of students has reached 13,000, and their annual income from English teaching will soon reach 27 million US dollars (Feng, 2007, p. 7). Another English teaching organisation, *Education First*, set up more than 70 branches in 28 provinces of China. The largest, however, is the *New Oriental School*, with more than 20 branches, 800,000 students, an income approaching 700 million Chinese Yuan and profits soon to be 12% yearly. The *New Oriental School* is in the process of expanding to 100 cities across China, and is likely to become “the *Wal Mart* in English teaching” (Feng, 2007, p. 7). A report released by EF believes that in the year 2013 alone, 30 billion *yuan* was spent in learning English in China, indicating that people are learning or would like to learn English in and out of school, and the demand is apparently driven by real, functional communication needs.

To summarise, Chinese governments at all levels have legitimized the importance of the English language in part to feed the nation’s modernisation



drive, and are trying to persuade people to learn it. Chinese people are witnessing an ‘English craze’ which is sweeping across Chinese kindergartens, primary schools, secondary schools and universities, emphasising a ‘no English, no future’ prospect. English tuition schools of all kinds are climbing on the bandwagon of this craze, including ‘Crazy English’ (J. Y. Li, 2009), and making considerable amounts of money (e.g., the home-grown New Oriental School). English is now seen as a vital communication resource, with powerful influence in international relationships, and it is now playing an important role both within Chinese communities and in international communications. “And this is only a tiny part of the vast and growing Chinese interface with English, likely to be one of the linguistic wonders of the new century” (McArthur, 2002, p. 2).

2.4.5 Research on attitudes towards CE

As discussed above, it is crucial to position ourselves, as users of this language, equally with inner circle users, rather than as followers always trying to measure up to their norms and standards. It is equally important, however, as a complementary concern, to become aware of attitudes towards other varieties of English. Timmis (2002) conducted a large-scale research study with a questionnaire (n=580, including 400 students from 14 countries and 180 teachers from 45 countries) and 15 interviews, on the question of their attitude



towards the NS and NNS model of English. For pronunciation, 67% of students preferred a NS accent. However, 64% of participants indicated a wish to “retain the accent of my country” (p. 242) at the same time. As these two percentages do not add up to 100%, there is obviously overlap between the respondents’ choices – something that Timmis (2002) did not seem to have indicated clearly.

A number of investigations have also been carried out specifically into CE, the focus of this thesis. For instance, Kirkpatrick and Xu (2002), X. Q. Hu (2004, 2005), He and Li (2009), D. C. S. Li (2009), He and Miller (2011), all investigate people’s views toward CE, using similar methods such as questionnaires and interviews. These six studies share another similar factor, too, in that they investigated Chinese people’s attitude toward CE since they are the real users and consumers of English. Results for a preferred model of learning English in these six studies all point to an inner circle English variety, as they tend to favor an NS-based pedagogical model in their learning. Yet, at the same time, results also indicate that respondents tended to agree to the inevitability of CE features. People’s shifting acceptance of CE as a legitimate, indigenised variety has also been tracked. He and Li (2009) conclude that:

...learners and users of ‘China English’ stand to gain if salient linguistic features of ‘China English’ are seen not as a source of embarrassment, but a



resource of empowerment, for recognising the legitimacy of ‘China English’ on a par with NS-based pedagogic models would help promote a stronger sense of ownership among users of English in Mainland China. (p. 85)

Through these studies, we can see that people in expanding circle countries such as China are still struggling to “bridge the gap between moral high grounds and down-to-earth concerns” (D. C. S. Li, 2009, p. 81).

2.5 Corpus linguistics and its use in research on English varieties

After reviewing topics and issues related to WE, this section focuses on the research design and methodology in this study.

2.5.1 Real language and the corpus revolution

There is nothing new about working with real language data. Linguists throughout history have always agreed on the importance of describing and understanding how living language is used in real life situations. Franz Boas (1858-1942), the pioneer figure in American linguistics, believed that when we investigate a language we need to describe the real, essential, everyday lexis, rather than that based on hypothetical expressions. He spent most of his lifetime



in Native American tribes collecting examples of their language, and finally produced *A Handbook of American Indian Languages* in 1911 (R. Q. Liu, 1995, p. 164). This prioritisation of real language in use heavily influenced Edward Sapir, who agreed that forms in use should be the focus of linguistic research. In 1921, he published *Language: An Introduction to the Study of Speech*, which suggested that when we study the forms of language, we have to focus on the lexis expressing the fundamental issues of the speakers' lives, and the way they are expressed. Ferdinand de Saussure (1857-1913), widely held to be the father of modern linguistics, shared this point of view. He also believed that, rather than to prescribe normative usage, the goal of linguistics should be to describe how language is actually used. These show that most schools of linguistics seem to focus on language in authentic use, despite differences in their theoretical orientations. In the 1960's, before the electronics revolution, and at a time when students did not have PCs on their desks, Leech (1966) foresaw the future of corpus linguistics:

...counting occurrences, in a large number of cases, is merely a laborious way of coming to conclusions one has already arrived at subjectively. More detailed quantitative analyses (requiring large corpuses and the aid of computers) can be expected to produce results beyond the insight of the native speaker. (Leech, 1966, p. 73)



So in the absence of computer corpora, Leech

spent most of his time watching TV—but it was strictly for business, not pleasure. He watched the commercial breaks from which to extract real English ad words and compiled his first corpus, which consisted of 617 television advertisements telecast in Britain in the early sixties and later on published his first book: *English in Advertising*. (cited in Thomas & Short, 1996, pp. 4-5)

During the same period, Jan Svartvik was involved in a milestone project, the *Survey of English Usage*, which had been led by Quirk since 1959. Again, it was not computer-aided. As Leech recalls:

In those days computers were rare, expensive, unreliable....I spent many late nights in Gordon Square to get inexpensive off-peak access to the Atlas machine, programmed by punched paper tape. (cited in Facchinetti, 2007, p. 17)

There were other obstacles, as Jan Svartvik relates:

In the unsupportive linguistic environment of the 1960s, being named a



‘corpus linguist’ only made you feel like discovering your name on the passengers’ list for the *Titanic*... To us on the Survey, surrounded by masses of authentic language data, both spoken and written, drawing the line between grammatical and ungrammatical sentences seemed a big problem. You will realise that our detailed analysis of corpora consisting of real-life language was very much swimming against the tide of the mainstream Chomskian view of language! (cited in Facchinett, 2007, p. 15)

Francis, who founded the first machine-readable Brown Corpus with Kučera in 1967, provides two other stories similar to that of Jan Svartvik above:

In the early 60s, when I was assembling the notorious Brown Corpus and others were using computers to make concordances of William Butler Yeats and other poets, one of my colleagues, a specialist in modern Irish literature, was heard to remark that anyone who would use a computer on good literature was nothing but a plumber. Some of my students responded by forming a linguistic plumbers’ union, the symbol of which was, of course, a monkey wrench. The husband of one of them, being a jewellery manufacturer, had a few tie clips made. (cited in Thomas & Short, 1996, p. 6)



And Francis wore this monkey wrench tie clip during the 1985 ICAME Conference at Lancaster University and said he would not part with it!

From all these early-day anecdotes about corpus linguistics, we can see that in the 1960's, anti-corpus attitudes were found both in linguistics and other circles. Of course, following the Brown Corpus, LOB, LLC and other more current and much bigger corpora like the BNC, COCA and MICASE were subsequently created as computer technology advanced. Sinclair (1991) describes the history of corpus linguistics in this way:

Thirty years ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular. (p. 1)

Ten years after that in 2001, Jan Svartvik wrote with title *Corpora are becoming mainstream* (Thomas & Short, 1996, p. 3), and that “It is hard to see a linguistic discipline not being able to profit from a corpus one way or another, both written and oral” (Cermak, 2003, p. 268). Wolfgang Teubert, editor of *The International Journal of Corpus Linguistics*, said that:



Today, the corpus is considered the default resource for almost anyone working in linguistics. No introspection can claim credence without verification through real language data. Corpus research has become a key element of almost all language study. (Teubert, 2005, p. 1)

How things have changed! The above shows that although authentic language in use has been the goal of many linguists, it was not until the last four or five decades that procedures for storing and investigating it became mainstream. There are some very good introductory guides to corpus linguistics, such as Kennedy (1998); Aijmer and Altenberg (1991); Thomas and Short (1996); Sinclair (1991, 2004), Hunston (2002); Meyer (2002); Teubert & Čermáková (2007); McEnery and Hardie (2012), Aijmer and Altenberg (2013), Friginal and Hardy (2014), among many others, although these are not reviewed here as they are beyond the scope of this thesis.

2.5.2 Frequency: key feature

One reason why corpus linguistics is revolutionary in contemporary language studies is that the standpoint is in direct contrast to the axioms of the then dominating Chomskyan approach to linguistics. The Chomskyan paradigm has held that linguists should be concerned with what is possible, with rules at the



highest possible levels of abstraction (universal grammar). Corpus linguistics differs in that it is only interested in what is probable, and probability of language production is only supported by authentic language data. Authenticity and frequency are more important than speaker intuition for corpus linguistics. So corpus linguists do not concern themselves with sentences like:

Colourless green ideas sleep furiously.

The rat the cat the dog chased ate died. (McIntosh, 1967, p. 188)

Nobody would ever use sentences like these in a real-life situation. Corpus linguistics makes possible the empirical analysis of language, and in so doing adds to its definition and description. It is research based on authentic data, in quantities appropriate to the purpose of the analysis, and its representativeness is achieved by good sampling.

2.5.3 Types of corpus

Corpora are built and used according to different standpoints. The following tables illustrate examples of different types of corpora:



- { General corpus: Brown, BNC (British National Corpus)
- { Specialized corpus: SCORE (Singapore Corpus of Research in Education)

- { Synchronic corpus: LIVAC (Linguistic Variation in Chinese Speech Communities)
- { Diachronic corpus: Helsinki Diachronic corpus (part of the Helsinki Corpus of English Texts)

- { Spoken corpus: SECCL (Spoken English Corpus of Chinese Learners)
- { Written corpus: WECCL (Written English Corpus of Chinese Learners)

- { Native speakers' corpus: LOCNESS (The Louvain Corpus of Native English Essays)
- { Learner corpus: ICLE (International Corpus of Learner English)

- { Monolingual corpus: corpus data only consists of one language(e.g., BNC)
- { Multilingual corpus: European Parliament Proceedings Parallel Corpus (versions of the same content in different languages)

- { World English variety corpus: ICE (International Corpus of English)
- { English as a lingua franca corpus: VOICE (Spoken corpus of English as a Lingua Franca)

It should be noted that one corpus may fall into more than one category, depending on the primary goals and perspectives. For example, the British National Corpus may be considered a general corpus, a native speakers' corpus and a World English variety corpus, and of course, it is a monolingual corpus.



2.5.4 Approaches in corpus linguistics research

Corpus linguistics itself is empirical, yet methods and approaches differ, depending on the way the data is collected and used in research. There are three main approaches:

- a. corpus-informed approach
- b. corpus-driven approach
- c. corpus-based approach

a. The corpus-informed approach was identified by Michael McCarthy (1998), which he believed to be suitable for textbook compiling. Results and data are analysed and adjusted according to the needs of students, providing empirical information as opposed to merely intuitive construction. Examples are Michael McCarthy's *English Collocations in Use* series, which consist of self-study materials and classroom activities (McCarthy & O'Dell, 2005, 2008).

b. A corpus-driven approach does not hypothesise anything. Rather, it summarises all the data and concludes anything supported by the evidence.

Essentially, this mode of research can be represented as follows:



—————→

Observation---hypothesis---generalisation---unification in theoretical statement

This approach is best applied to the description of language, free from the bias of the researchers' personal interest and experience.

- c. A corpus-based approach is very much about hypothesis testing. It embodies a dichotomy methodology in linguistics, namely the testing of hypotheses with exploration and description (Gilquin & Gries, 2009), using data as evidence to support or negate research hypotheses that are in the form of research questions.

This inclusive approach utilises the corpus, but does not attempt to reject or disprove other approaches that might be used in the field of linguistics. Through this kind of research, some traditionally accepted concepts and theories can be further tested or extended. Corpus-based research is characterised by the use of probability in terms of frequency and statistics, as summed up in the corpus data. For example, three function word types (*the, of, and*) account for about 11 percent of all the English words we ever hear, speak, read or write (Kennedy, 2003, p. 84), thus their probability of occurrence is higher than any other lexical item. Their combinations, such as *a lot of, lots of* are more likely to occur than, say, *too many*. The present research is therefore clearly an example of a

corpus-based study. The support or negation of the research hypothesis will depend on the results derived from the data analysis: the data will speak for itself.

The corpus-based approach may be regarded as the application of statistics and empirical study in linguistics, so the general steps involved in the research procedures are similar to those of most empirical studies, as follows:

- a. put forward research hypothesis
- b. follow one classification of terms and operationalisation
- c. choose suitable corpora
- d. choose suitable software for analysing the corpus data
- e. retrieve data from corpora according to criteria set in b
- f. deal with statistics of the data
- g. discuss and explain the statistics
- h. draw sensible conclusions from the research questions as well as the hypotheses

(Liang, Xu & Li, 2010, pp. 178-181).

In chapters 4 and 5, I will follow this framework and illustrate in detail the corresponding data sets to address each of the research questions.



2.6 Research on features of CE

Certain lexical features distinctive to Chinese speakers of English are pervasive in Mainland China. Wen (2006) searched the *China National Knowledge Infrastructure* (CNKI) (Online journals) and Digital Dissertation Database and found that there are 162 corpus-based studies on features of Chinese speakers of English, 91 of which are on lexical features. The main focus of this thesis, therefore, will also be on lexical features, specifically on lexical collocation, which is relevant to characteristics of English varieties. However, before proceeding with describing the methodological design in this research, an outline of previous work done on CE features is in order.

The book *Chinese English* (Xu, 2010) considers the development of CE within the World Englishes paradigm, and says a great deal about the extraordinary development and spread of English in general, in particular in China. The work not only “makes a unique contribution to the field” but also enable us to “sensibly talk about a Chinese variety on English” (Kirkpatrick, 2010, p.1). Xu uses as a research source data from CE interviews, newspapers and short stories and found that there is ‘Chineseness’ reflected in Chinese English lexis, meaning that readers or listeners must call upon Chinese contextual knowledge in order to fully understand it.



The major findings of his investigation indicate that CE lexis comprises three groups: Chinese loanwords (e.g., *guanxi*, *pipa*, *renminbi*); nativised Chinese English words (e.g., *family planning*; *backdoor*) and words that are common to other varieties of English (e.g., *table*, *car*). The syntactic features revealed include: an adjacent default tense, null-subject/object utterances, the co-occurrence of connective pairs, and distinctive sequencing of modifier and modified in nominalisation. He states that these features represent a transfer from Chinese into English by CE speakers. Xu also illustrates findings of discourse features of Chinese English: patterns, quotes, metaphors, proverbs and address terms that have encodings which reflect Chinese pragmatic norms and cultural values. Xu's (2010) empirically-grounded study helps us to see characteristics of Chinese English at a number of levels, and to ascertain that Chinese English is likely to become a variety of English, and that it is important for the international English-speaking community to become familiar with this expanding circle variety of English. However, the author admits that the body of data is relatively small, and more research should be carried out on more and wider data sources. The literature reviewed in section 2.3.1 of this thesis, featuring articles published in the 1980's and 1990's, all provide examples as evidence for their arguments. However, those examples of China English were generated by the authors, rather than derived from authentic language as used by real-life CE speakers/writers.

Xu's research differs in that it is supported by real language data, and thus is convincing and remarkable in its own right. 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 frequently in English. Because of the society he grew up in and the education he received, he is a good example of an expert China English user. The evidence of the CE found there is thus very much a focused case study.

Sun (2011) focuses on degree adverbs in Hong Kong English by comparing them to Singapore English. Her investigation of ICE-Hong Kong (HKE) and ICE-Singapore (SE) found that although in general HKE and SE are similar to British English (BrE) in their use of degree adverbs (e.g., *completely*, *totally*, *absolutely*), they display some commonalities that distinguish them from BrE. In addition, she found that there are also identifiable differences between the Asian Englishes, with SE exhibiting more variety-specific features than HKE. Examples are:



Table 2.9: Features of SE (as contrasted to HKE) (Sun, 2011, p. 201)

Feature	e.g.
Overused	<i>very well, very nice, very bad, quite ok/alright, so bad, so terrible</i>
Underused	<i>really good, really nice, extremely difficult, particularly important</i>
common in HKE and BrE, absent from SE	<i>absolutely right, really difficult</i>
Colligational preference in SE than HKE	<i>so very + adj/adv</i>

Sun (2011) claims that all these differences are manifested in the form of frequency of use. In her observation, four kinds of frequency differences are distinguished:

- a) the overall frequency of degree adverbs in each variety
- b) their frequency in different genres
- c) the frequency of individual adverbs
- d) the frequency of a number of commonly used collocations (p. 197).

The differences may be related to overt differences in the respective positions of English in these two societies and the resulting stage reached by their varieties of English in the evolutionary cycle of New Englishes (Schneider, 2003). SE displays more variety-specific features than HKE, which confirms

the claim made by Mukherjee and Gries (2009) that the more advanced a variety is in the evolutionary cycle, the more divergent it is from its ‘parent’ variety. So Sun’s (2011) study provides further empirical evidence for the position of HKE and SE in the evolutionary cycle of New Englishes. The findings in her thesis suggest that structural nativisation has emerged in HKE and SE with respect to degree adverbs. Through these investigations and findings, she also argues that Chinese cultural values could play a role in some characteristic usage of degree adverbs in HKE and SE. Generally, the importance of the study lies in the fact that it corroborates the idea that an investigation of forms and structures belonging to the common core of English grammar can shed light on the status of the nativisation and acculturation of a New English.

Chen (2010) investigates the discourse-grammatical features of advanced Chinese learners of English by comparing them to those of British students. She takes a NNS versus NS perspective on the English used by the Chinese learners, using the LINDSEI and LOCNEC corpora. Frequency and distribution of different grammatical forms for theme, person reference and tense aspect are analysed using corpus tools. Interestingly, despite Chen’s claim that the goal of her investigation is not to codify features of China English, some significant features did emerge through her investigation:



- a) Marked theme constructions
- b) Topical constructions for personal reference
- c) Non-standard and base forms of verbs
- d) Aspect marking (Chen, 2010, p. 225).

Chen's results show distinct tendencies in these four categories, which might be considered features of CE. Furthermore, some of the tendencies are "found to be consistent regardless of variation in task type and proficiency level" (p. 227). Her findings suggest that marked choices in native speaker speech (such as marked theme and full nominal reference in the context of textual continuation) are used more frequently in L2 speech as unmarked forms of discourse development. Additionally, some textual and pragmatic functions of these forms in NS speech are dropped, and replaced by others in L2 speech, which may be due to language transfer and/or communicative strategies used by learners to cope with the anxiety or tension created in producing spontaneous speech. Since Chen is taking a NS versus learner perspective, she suggests that the deviant patterns found are influenced by the syntactic and discourse patterns of L1, by performance constraints, and by inadequate teaching and learning of the target pattern, which she predicts will cause persistent difficulty for learners of English in China. Thus she advocates that the gap between grammar and discourse should be filled by clearer teaching, and that English teaching in



China should move more towards teaching discourse-grammatical features, by teaching grammar as a discourse phenomenon. Chen's study is thought-provoking in that she uses homogenous corpora – learner corpus data collected from students of comparable profiles in Europe and China – to compare the features targeted for analysis in her research proposal. It is also a convincing account of the language of Chinese learners of English, since it is based on the usage of university students. However, her concern is with deviations and learner errors, and the focus of this thesis is not on learner language weaknesses, but on the presence of other, more substantial features which may contribute to the reassessment of CE as an emerging new variety.

There are also studies that take a WE perspective towards CE features. For example, C. Gao (2008) and Yu (2009) investigate CE features in verbs and adjectives. Yu compares a corpus of China English News (written) with the BNC news sub-corpus, focussing on adjectives. The target of her study is the high-frequency usage of adjectives, along with collocations of adjectives, in these two varieties of English. For example, the adjective *great* collocates more with the following words in CE: *change(s)*, *achievement(s)*, *contribution(s)*. Gao's investigation takes a similar approach to Yu's with regard to research methodology and data analysis, but the difference lies in the focus: Gao studies the use of creation and transformation verbs in China's English newspaper (e.g.,

DEVELOP), using a corpus of British newspapers as its reference.

Unfortunately, much of the previous research on features of CE is deficient in a number of ways. However, studies of lexis in CE have been limited to word-class-based searches for features such as verbs, adjectives and adverbs. Little attention has been paid until now to the question of the deeper layers of colligational features, and still less to the culture-specific semantic preferences involved in collocations.

To sum up, in this chapter, we have discussed the situation of English learning and teaching in China. We have seen that:

- a) English is important to students ranging from primary three to university/post-graduate level, because it is a major school subject, and they have to pass English proficiency tests at all levels at different stages, such as for entry to and exit from higher education, or for specific jobs, typically in the white-collar workplace.
- b) English learning has been a key part of curriculum planning at a national level since the 1980's.
- c) The number of English users/learners/speakers in China has increased exponentially, reaching hundreds of millions of people in the past 30 years.



They have been taught mainly in schools by Chinese speakers of English.

- d) The role of English is more important than before, and the attitude of Chinese people towards English and English learning is the most positive in history.
- e) Research into features of Chinese English is still scarce.

We can also say, in a rough-and-ready way, the relatively brief progress of CE research can be seen as falling into three stages:

- a) the existence of CE and the issue of terminology;
- b) attitudes towards CE; and
- c) linguistic features of CE.

This thesis is in line with He and D. C. S. Li's (2009) point of view that:

...there is some evidence that 'China English' is gradually emerging, following its natural path of development, although it is quite impossible to list all the linguistic features of 'China English' exhaustively at the moment for several reasons, such as insufficient research. Therefore, more research is needed to identify salient linguistic features of 'China English' as found in the popular usage patterns of the majority of



speakers and writers of ‘China English’, in both formal and informal contexts of social interaction. (2009, p. 74)

Within sights obtained from the above literature as the background, this thesis would like to move on to investigate aspects of the CE phenomenon in a more detailed and, where appropriate, quantitative manner, and to raise questions that seem to us to warrant future attention.

2.7 Objectives and research questions

Despite a large amount of research on the lexis of CE that has been carried out, there is an urgent need to extend the focus of research to include longer word groups or strings of lexical items. With this research focus and objective in mind, the thrust of the whole thesis is to present empirical evidence to account for some of the ways in which English is used by Chinese people (Mainland China, as justified in Section 1.1). It explores the collocational features of China English (CE) from the perspectives of World Englishes (WE) and English as a lingua franca (ELF), thus viewing CE in a different light compared with that from the perspective of Error Analysis (EA).



The overall claim of this thesis is that whereas lexical innovations which represent local meanings of socio-cultural import constitute a significant type of evidence and argument when assessing to what extent a new variety of English is emerging, CE-specific collocational patterns which constitute an important type of collocational or lexical innovation represent another important body of evidence in regard to the process whereby CE is evolving into a new variety of English.

The theory behind the present account is known as the theory of collocation (Sinclair, 2004, p. 39). This thesis proposes that there are distinct CE features in terms of collocation, and the corpus investigation results will provide empirical evidence to substantiate this claim.

The purpose of this thesis is twofold. Firstly, it aims to demonstrate how collocational features of CE may be observed and described with the help of large corpus data, and secondly to explain why these features are linked so distinctively to Chinese speakers/writers of English. However, this second goal should not distract readers from the main goal of this thesis, namely to identify and demonstrate the distinctive collocational features of CE in ELF communication, which may eventually make possible the assessment of the extent to which CE is evolving as an autonomous variety in its own right. It is

hoped that the findings of this study will make some contribution to the linguistic description of CE. In light of the foregoing goals, the specific objectives of the research are threefold:

1. To provide a descriptive account of the use of CE in terms of collocation;
2. To examine whether the use of these particular collocational features exhibits any distinctive characteristics marking the different linguistic backgrounds and worldviews of Chinese users of English in general; and
3. To explore the possibility of advancing a theoretical framework for researching CE collocational features.

Focusing on collocational patterns of high-frequency general nouns, this thesis addresses the following research questions (RQ's):

1. Are there distinctive collocational features in the English used by Chinese?
2. Are there colligational preferences in CE?
3. Are there semantic and prosodic preferences in CE?
4. Is it viable to propose a framework for collocational features of China English based on the informed answers to RQ's 1-3?



2.8 Summary of chapter

This chapter has mainly reviewed six aspects related to the research topic. Firstly, topics related to WE perspective have been reviewed: models of English, contradictory views towards norms and standards of English, and the pluricentric approach to describing and analysing language variation. Then, the interface between these WE standpoints and CE was probed, including the issue of variability and linguistic prejudice. After that, this chapter delineated the definition of China English, English learning and teaching in China, and research findings of CE to date. We have also reviewed key issues in corpus linguistics, which provides the basis of and rationale for the methodological framework (chapter 3) and procedures of data analysis (chapter 4).



CHAPTER 3

METHODOLOGY AND THEORETICAL RATIONALE

3.1 Introduction

The rationale for this corpus study of collocational features in CE will be explained in this chapter. The diversity and richness of collocation studies could hardly be captured in one summary; the organization of this part is therefore centred around some of the most relevant issues only. Firstly, an account of the concept of collocation will be provided. Then, the research methodology adopted in this study will be elaborated before presenting the specific processes involved for identifying CE collocational patterns.

3.2 Collocation theory

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. Uncovering the extent of word associations and how they are manifested in collocations has been an important area of study in corpus



linguistics since the 1960s. Consider these examples:

strong tea	*powerful tea
black coffee	*dark coffee

Although the examples in the right-hand column present no grammatical problems, those in the left-hand column are preferred by most speakers. Here the co-occurrence of words *strong* and *tea*; *black* and *coffee* is referred to as one example of ‘collocation’ (Halliday, 1966, p. 149). Many apparent synonyms, such as *strong* and *powerful*, can have characteristically different collocations. In its non-technical sense, collocation is defined as:

a term used to describe 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, p. 62)

As a technical term in linguistics, however, collocation implies rather more than mere co-occurrence side by side. This phenomenon of repetition of word



patterns has always been one of the central concerns of linguists over time.

Below is a short sketch of the history of collocation studies.

According to Kennedy (2003, p. 478), Palmer first adopted the term ‘collocation’ to refer to recurring groups of words, and further defined it as “a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts” (Palmer, 1933, p. i), an example of which might be *on the whole*. Palmer discussed collocation from the perspective of English language teaching, and his view was paralleled in different branches of the language sciences.

Among a number of scholars who took account of the phenomenon of collocation, J. R. Firth was the most often cited. His famous maxim “*You shall know a word by the company it keeps*” (1968, p. 179) emphasises the importance of both linguistic and situational contexts for the description of language. What is most revolutionary in Firth’s argument is that the habitual collocations in which a word appears are part of that word’s meaning and are a type of “mutual expectancy between words” (Firth, 1968, p. 181). This could also be interpreted as: collocating words predict one another, in the sense that where we find one, we can expect to find the other. He uses the words *dark* and *night* as an example of his argument: *dark* is characteristically used in



conjunction with *night*; collocability with *night* is one of the meanings of *dark* (Firth, 1957, p. 196). The meaning here is simply a characterisation of the “other word-material” (Firth, 1968, p. 180) with which *dark* is often used. Sinclair, whose work draws on the Firthian tradition, is one of the key pioneers and contributors in this field.

One of Sinclair’s insights concerning the significance of the concept of collocation is that “[w]ords which stand in such a relationship can be said to predict one another because the presence of one makes the presence of the other more likely than it would otherwise be” (Sinclair, 1966, pp. 417-418). Furthermore, “[w]ords are ‘collocates’ of each other if, in a given sample of language, they are found together more often than their individual frequencies would predict” (Jones & Sinclair, 1974, p. 19). One of Sinclair’s best known contributions to the field of collocation research is his work using corpora and corpus linguistic techniques to demonstrate what had been proposed by those early linguists about collocation.

Before Sinclair, linguists sought examples to argue for the existence of collocations of all kinds by resorting to personal intuition or observations in daily lives. With the help of corpus research methods, however, large numbers of examples of collocation patterns are found empirically. The logic behind this



is summarised in his famous saying “the language looks rather different when you look at a lot of it at once” (Sinclair, 1991, p. 100). With the technological aid of computers that can store a billion-word-token corpus, and tools that can analyse it, it becomes possible to observe patterns of language use which would otherwise be invisible, even given thousands of years of textual study.

Further to advocating a ‘corpus revolution’, Sinclair further developed the topic of collocation. Based on the Birmingham Collection of English Text (also known as the *Bank of English*) which represents British English, Renouf and Sinclair (1991) put forward another significant case for collocation research: collocational frameworks. One of the examples they used is the pattern:

$$\frac{a + ? + of}{}$$

e.g., *a lot of*, *a couple of*, *a kind of*, *a matter of*

They argued that the pattern *a + ? + of* (and the like) is not accounted for in descriptions of the language:

In grammar, it lies somewhere between word and group; in lexis, it is missed by conventional definitions of collocation, lexical item and phrase;



semantics has no means of dealing with such grammatical co-occurrences, either as two individual items or as a unit with interdependent meaning. (Renouf & Sinclair, 1991, p. 129)

Through corpus-derived statistics, they demonstrated that the two very common grammatical words, one on either side, offer a firm basis for studying collocations. The significance of this investigation is that collocations are not necessarily continuous. They articulate this observation as follows: “Collocational frameworks in English consist of a discontinuous sequence of two words, positioned at one word removed from each other; they are therefore not grammatically self-standing; their well-formedness is dependent on what intervenes” (Renouf & Sinclair, 1991, p. 128).

Up to this point, Sinclair’s linguistic theory of collocation was mainly concerned with finding more sensible ways to present and explain language patterning. By the time he wrote his later monograph *Trust the Text* (2004), however, Sinclair has taken these ideas much further. He then summarises decades of work on collocation by setting four parameters for the ‘meaning’ of any given specific lexical item as outlined in Table 3.1:

Table 3.1: Four parameters of meaning (based on Sinclair, 2004, p. 174)



collocation	the co-occurrence of words (e.g., <i>dark night</i>)
colligation	the co-occurrence of words with grammatical choices (e.g., adjective + noun)
semantic preference	the co-occurrence of words with semantic choices (e.g., typical subjects of a verb)
semantic prosody	the expression of attitudinal connotations of a word (something being desirable/undesirable)

These four parameters form what will be referred to in this thesis as Sinclair's theory of collocation, upon which the present investigation is based.

It should be noted that other scholars have responded to these four concepts with varied interpretations, particularly the last one. For example, Stubbs (2009) in his memorial article to Sinclair interprets semantic prosody as 'communicative purpose' (p. 125). Theoretical deliberation on how to define semantic prosody continued but its conceptual delineation remained unresolved. This research follows Sinclair's interpretation: "Many uses of words and phrases show a tendency to occur in a certain semantic environment" (Sinclair, 1987, p. 321). As a corollary, such items may come to carry "an aura of meaning that is subliminal, in that we only become aware of it when we see a large number of typical instances together" (Sinclair, 2004, p. 18). One example

is the word *happen*, which, according to Sinclair, characteristically appears together with “something nasty that has happened or is going to happen” (p.18). Another is *set in*, which again collocates with “nasty things like bad weather” (p. 18). As Partington remarks, “a phrase like *good times set in* would be highly marked” (1998, p. 67).

So this thesis uses three of the original definitions of the terms in this theory, while reinterpreting the last, as shown in the following list:

Collocation: “the co-occurrence of words with no more than four intervening words” (Sinclair, 2004, p. 141)

Colligation: “the co-occurrence of grammatical choices”
(Sinclair, 2004, p. 141)

Semantic: “the restriction of regular co-occurrence to items which share a preference semantic feature” (Sinclair, 2004, p. 142)

Semantic: the driving force behind the selection of the core and the other prosody co-selections comprising the lexical item. Semantic prosody of a word is established according to their attitudinal bias, i.e. on the basis of whether they have a positive, negative or neutral association.

In sum, there are two key terms underlying the theoretical rationale in this study, namely ‘co-occurrence’ and ‘high frequency’:

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 our knowledge of these words or concepts. The semantic profile of one word is established by the co-occurrence of other words.

High frequency: the number of occurrences of a pattern or co-occurrence is sufficiently high to qualify them as candidates for being ‘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. In every case, justification of the cut-off point is usually provided. The same practice will be followed in this thesis.

In the rest of this thesis, features found in these four structural subcategories will be referred to as ‘collocational features’.

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’ (Parley & Syder, 1983), ‘chunk’ (Sinclair, 1991), ‘lexical phrase’ (Nattinger & De Carrico, 1992), ‘lexical bundle’ (Biber, Johansson, Leech, Conrad & Finegan, 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 thesis, all these terms are treated as synonymous with one another, although the word ‘collocation’ will be used throughout.

3.3 Research methodology: a corpus-based approach

As shown above, collocation theory attempts to formulate rules and constraints for the co-occurrence of different properties or levels of words. This section will discuss the reason why a corpus approach is chosen to investigate the given research questions.

3.3.1 What is a Corpus

In modern linguistics, a corpus is usually defined as a body of naturally occurring language, although different definitions abound. Some examples follow:



- a. “[C]omputer corpora are rarely haphazard collections of textual material. They are generally assembled with particular purposes in mind, and are often assembled to be (informally speaking) representative of some language or text type.” (Leech, 1992, p. 116)
- b. “A corpus is different from a random collection of texts or an archive whose components are unlikely to have been assembled with such goals in mind. The term corpus as used in modern linguistics can best be defined as a collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information.” (McEnery, Xiao & Tono, 2006, p. 4)
- c. “In the language sciences a corpus (plural corpora) is a body of written text or transcribed speech which can serve as a basis for linguistic analysis and description. The complication and analysis of corpora stored in computerized databases has led to a new scholarly enterprise known as corpus linguistics”. (Kennedy, 1998, p. 1)

Linguists from different persuasions and camps are slightly divided over how corpus and corpus linguistics should be defined. It is my view that language-focused research should be corpus-driven without being restricted by it, so in this thesis I would like to adopt a broader definition of ‘corpus’ as follows:



A corpus is a large collection of a) authentic texts representative of a particular language variety (spoken and written), which is b) machine-readable so that large numbers of texts can be analysed and examined repeatedly and iteratively.

3.3.2 Why use a computer to study language

In 2.5.1 we saw that real language use has been a shared focus for many linguists over a long period of time. Boas, for example, spent most of his lifetime among American-Indian tribes collecting real Indian languages; his efforts culminated in the seminal *A Handbook of American Indian Languages* in 1911. This might have been different if he had had a computer. Corpus data can be dealt with by computer tools in a quantity and at a hitherto unthinkable speed. Software-readability is the most important attribute of modern corpora. Corpora are in fact typically computerised bodies of language, and “computer corpus linguistics” has been suggested as an alternative name for corpus linguistics (Leech, 1992, p. 106). Electronic corpora have advantages unavailable to their paper-based equivalents, when one considers the speed of processing they afford, and the ease with which they can manipulate data. Examples can be searched for rapidly, and retried any number of times, at minimal cost. As Tognini-Bonelli (2001) observes, the computer has affected the methodological



framework of linguistic enquiry (p. 210). In section 2.5.2 some of the key features of corpus linguistics were discussed. However, one further, notable, feature of corpus-based research is the quantity of data that the computer and its software tools can deal with. The following table shows the sizes of some of the most popular corpora currently available to linguists for research purposes:

Table 3.2: Examples of popular corpora and their size

Name	Full name	Size (in word tokens)
BNC	British National Corpus	100 million
OEC	Oxford English Corpus	2 billion words
BoE	Bank of English	450 million words
CIC	Cambridge International Corpus	1 billion words

*the sizes of the corpora are based on Baker (2010, p. 12)

These large quantities of language data are available for computer users thanks to the tools created to access them. Just as scientists require the use of specific technology in order to examine the human body, corpus linguistic also requires tools in order to conduct corpus analysis. To put it in another way, a corpus is not particularly valuable unless it is used in conjunction with computer software that can quickly and accurately perform manipulations on its contents. Some corpora come with their own built-in tools, while other tools function as

stand-alone platforms, capable of carrying out analysis on any text or corpus which the user specifies. Some tools are web-based, allowing users to search the corpus data online, while others need to be run from the users' own PC. Corpora are simply linguistic data unless specialized software tools are used to view and analyse them. Hence, corpus tools are becoming an important factor as corpora become larger and the statistical analysis of linguistic data becomes more complex. Table 3.3 shows a short list of corpus tools currently available:

Table 3.3: Popular corpus tools

Name	Creator	Type
WordSmith Tools	Mike Scott	stand-alone
AntConc	Laurence Anthony	stand-alone
Xaira	Lou Burnard	stand-alone
Sketch Engine	Adam Kilgarriff	web-based
View	Mark Davies	web-based

It is the corpus tools that help find the linguistic evidence which researchers hope to use to address their research questions, as mentioned above, in a repeatable and iterative manner. Many corpora are homemade, built by the researchers themselves from material they have collected, stored and then analysed with concordancing or other corpus tools. “There is no reason why we

shouldn't use corpus techniques on smaller texts" as long as researchers do not "try to make generalisations about language use beyond that particular text...any findings we make are unlikely to be representative of all language use" (Baker, 2010, p. 7). In Chapter 2, it was pointed out that CE is seen as a developing variety in the expanding circle of the family of WE, and it could be safely argued that researching features of CE falls within the scope of sociolinguistics. Baker (2010) states that:

A useful way in which corpus approaches can aid sociolinguistics is in providing large amounts of existing data (or standards about techniques for building representative corpora of a population), along with computational tools and procedures which allow common (and rare) language patterns and frequencies to be identified quickly and accurately and compared across different populations. (p. 9)

The advantage of using a corpus to research into a language variety is that:

Compared with the more traditional intuition-based approach, which rejected or largely ignored corpus data, the corpus-based approach can offer the linguist improved reliability because it does not go to the extreme of rejecting intuition while attaching importance to empirical



data. The key to using corpus data is to find the balance between the use of corpus data and the use of one's intuition. (McEnery, Xiao & Tono, 2006, p. 7)

Accordingly, the corpus-based methodology is not mutually exclusive with an intuition-based approach. The two are complementary and must be so if as broad a range of research questions as possible are to be addressed by linguistics.

In summary, some basic issues of this research have been addressed in this chapter, namely its theoretical basis and the rationale behind its methodological framework. In the following chapter, methodological details from preparatory steps and procedures of investigation to data-analysis and retrieval will be described.



CHAPTER 4

RESEARCH DESIGN

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. Designing the procedures to spot out such patterns constitutes the core issue of the investigation. This part of the thesis presents relevant research designs that lead to the goal as set out in chapter one. A framework of the steps taken in the investigation process of this research is thus explained in the following section.

4.1 Research framework

The research design is chiefly based on Sinclair’s paradigm for researching patterns of collocations (2004, p.39):

collocation → *colligation* → *semantic preference* → *semantic prosody*

This research paradigm involves taking a number of steps in the investigation of



co-selection of words (Sinclair, 1987, 1996, 1998, 2004) which can eventually lead to the identification and description of collocational variations. These features are significant with regard to a number of categories. Some of these collocational features are overt, whilst others are less obvious. Figure 4.1 is a schematic representation of the theoretical framework adopted in this thesis:

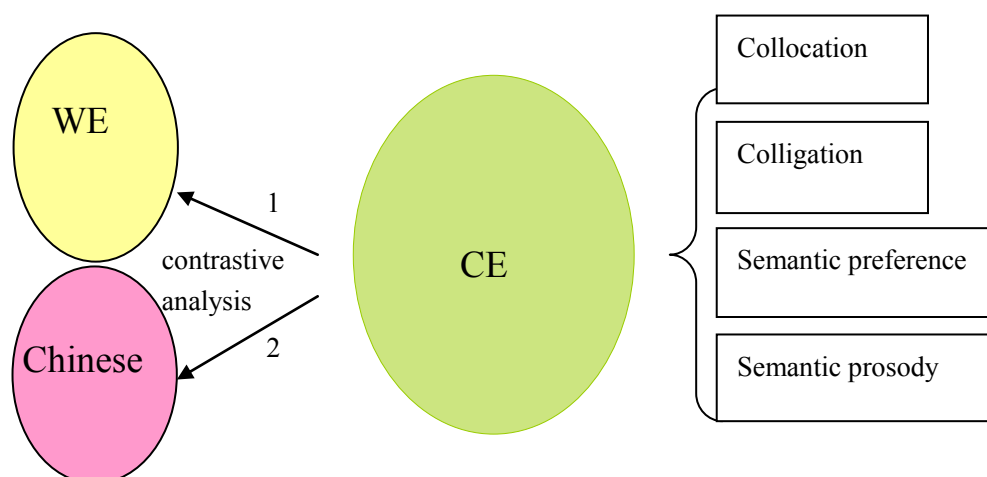


Figure 4.1 : Research framework

4.2 Data

This section of the thesis offers details concerning the data used. Since this is a contrastive study, each pair of datasets for comparison is presented in turn. The target corpus data consists of 37 million words of English collected from Mainland Chinese users and is referred to as CEC. British National Corpus BNCweb (CQP-edition) (referred to as BNC) is used as the reference corpus (see Table 4.1):

Table 4.1: Target corpus and reference corpus

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

4.2.1 CEC

CEC (W. Z. Li, 2011) is the largest corpus currently available that represents the English used by educated Chinese people (mainly people receiving undergraduate level of education or above). It is large enough (37,470,040 word tokens) to generate representative results. This corpus was constructed explicitly for the purpose of investigating CE linguistic features. To this end, the data collected seek to meet the criteria of genre balance (e.g., written events

include educational, leisure, natural and social sciences, and business studies and communication, etc.) and a range of domains (see Table 4.2 below). The writers of texts in CEC share similar linguistic backgrounds in Mainland China.

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

For example, CEC includes extracts from regional and national newspapers published in Mainland China, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published information bulletins, published theses and papers, among many other kinds of written text. The comparable nature of these two corpora in terms of domain classification is listed in Table 4.2 below:



Table 4.2: Domains of CEC and BNC

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

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

Written English represents one consistent type of English output of educated CE users; for obvious reasons, written data, which is easier to obtain in useful quantities than spoken data, serves as a convenient, ready-to-use database for corpus-based linguistic investigation. The readership of these written English data is two-fold: first, those Mainland readers who have the ability and will to access information disseminated in multiple channels and languages, including English; second, readers who are driven by a desire to improve their English by giving themselves additional opportunities and exposure to that target language more or less regularly. Foreigners with no Chinese language background could

only access information (e.g., through English newspapers in China) or through web pages whose contents, in the Chinese context, are produced by CE users. Today, there is an increasing number of educated people in China who would like to read a variety of texts written in English. Therefore, there is a need for the written sources to be comprehensible and to convey practical information with the two kinds of readership – Chinese and non-Chinese – as target consumers, which is clearly the key function that these home-grown written English outputs are serving. Produced with that purposeful goal and target readers in mind, the components of CEC are thus ideal sources for the current investigation.

As for the representativeness of CEC data, this thesis follows the same premises that previous pioneering studies adhered to in regard to the representativeness of corpus data. Leech (2007), for example, argues that “there is a scale of representativity” (p. 144) in data. Similarly, McEnery and Hardie (2012) state that “the measures of balance and representativeness are matters of degree.” (p. 10). Thus the representativeness of CEC should be calibrated by the extent to which it truly reflects the preferred linguistic features of Mainland Chinese users of English collectively as a whole. Texts in CEC represent CE in two ways: ‘mainstream’ and ‘general’. CEC represents mainstream CE, because it includes articles from mainstream newspapers which are categorised into ten



components, whose sizes and sub-categories are listed in Table 4.3. The texts were sampled from different genres in more than 30 domains, including journal articles, academic essays (wide coverage of topics/disciplines), news reports (reportages and reviews), editorials, public relations documents from public and private organisations, etc. These genres are all comparable to those in BNC.

Table 4.3: Domains and description of the CEC data

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

*Note: tokens and types are calculated with the help of the corpus tool AntConc (Anthony, 2011).

Firstly, using AntCoc, the names of authors of articles included were checked to establish whether they are Chinese. Chinese names follow a format (e.g., ZHANG Kexin), while names of people not from China would look different (e.g., Matsuda Aya, which is not recognisable by the *pinyin* system). This precautionary check is to ensure that all data represent CE.

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

a. Representativeness in terms of size

There are 37,470,040 word tokens in the CEC. Information about frequency is one of the most obvious benefits that a corpus can provide, which cannot be provided by any other mode of linguistic analysis. Clearly, then, in a corpus of the size of the CEC, the information about frequency is more convincing than intuition-based theorising. The importance of frequency information is also endorsed by the mathematical theory which calculates words (or lexemes) according to the patterns in which they appear, and generates hypotheses from it. The mathematical theory in support of this method is Zipf's Law of word distributions. Zipf (1935,

1949) held that in any language corpus, the frequency of any word is inversely proportional to its rank in the ranking table of frequency, and the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. Zipf's Law gives a more heavily weighted importance to the most frequent words than would be expected according to normal distribution in language. In other words, the larger the data, the more suitable it is in terms of obtaining reliable and representative frequency information.

b. Authenticity in terms of function

As discussed earlier, corpus linguistics focuses on authenticity. Some may argue that written articles are not as authentic as daily face-to-face conversation. However, as indicated in Chapter 2, at present there are millions of competent users of English in China (as a result of the education they received, and the computer networks that have evolved as platforms with which CE users communicate with the rest of the world). However, English is not commonly used between Mainland Chinese when there are no English-speaking people around, so the collection of written and spoken data from spontaneous situations is not only difficult but also not quite feasible, at least not at present. In this light, CEC could be seen as representing the way in which Chinese users of English write when



conveying information to target readers they hope to reach, at home or abroad. The authors who wrote the articles in question may have overseas experience and speak ‘native-like’ English; however, even here, CE features will be found, if they are indeed CE features. The ultimate purpose that each article fulfills is to accomplish the communicative function in written form, which is authentic too in every sense of the term.

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

4.2.2 BNC

The British National Corpus (BNC) (Leech, Rayson, & Wilson, 2001) is a 100-million-word structured collection of spoken and written texts. The corpus was compiled by a consortium of universities, publishers and the British government in the 1990s, to be representative of the spoken and written English used by British people toward the end of the 20th century, with written data amounting to around 90,000,000 tokens. The front-end interface of the BNC is



available on the BNC_web (<http://bncweb.lancs.ac.uk/>). After registration, users can access the BNC_web according to their search targets. The search in the case of this thesis was limited to written form in the BNC_web, in order to ensure that it is comparable in size and genre to the CEC.

There is another well-established corpus of English representing another influential inner circle variety — COCA. COCA is a large-scale corpus of American English, which is another reference corpus for comparison and contrast with CE. However, as was mentioned in Section 4.2.1, CEC was conceived and built according to the framework of BNC, with comparable topics, genres and sizes of each domain by design. That makes BNC a more suitable reference corpus for the purpose of the current study. Further comparison and contrast with COCA-derived data will no doubt be a promising direction of future research.

4.3 Preparation

Some pre-analysis modifications were made to the data. Firstly, the texts were cleaned by going through the following procedures:

- a. The files were joined, using a programme called *Filejoin* (a freeware,



anonymous).

- b. The files were then cleaned, and unnecessary spaces or blank lines etc., were deleted, using *Text-organizer* (Fenglin, 2005).

They were then stored, and are ready for use. At this stage, the CE data is raw data without annotation or tagging.

4.4 Procedures

When asked whether Chinese speakers of English employ distinct features of their own, people from English-speaking countries would probably respond “Of course, they do!”. Before working on this research project, the author asked 10 people (all university graduates) from inner circle countries randomly, and each of them gave similar answers, some even laughing at the question because the answer is so obvious and the question so naïve. However, it should be kept in mind that language is about meaning-making, so when millions of people tend to share some kind of ‘colour’ in their meaning-making activities, it is worth conceptualizing what these ‘features’ can tell us about their speakers/writers and how.

The challenge that faces linguists in attempting to describe linguistic features in terms of collocations in the English produced by Mainland Chinese people is



how we identify them and argue that they are unique to them. The present study proposes that one way to answer this question is by conducting contrastive analysis. Specifically, it aims to compare the English used by Chinese speakers and that used by speakers of other varieties of English, so that we can see the differences and, wherever possible, compare these CE features with the corresponding features in Chinese—the L1 of CE users. The first comparison tries to answer the question, “What is different?” while the second deals with the question “Why might this be unique to Chinese?”.

The first step towards identifying and describing CE collocational features is to sample the data for analysis, trying to determine and narrow down the specific targets. This step serves as a ‘pilot’, the outcome of which would inform and help narrow down the scope of the main study. So at this point, the research method is a qualitative one and the procedure may be outlined schematically as in Figure 4.2.

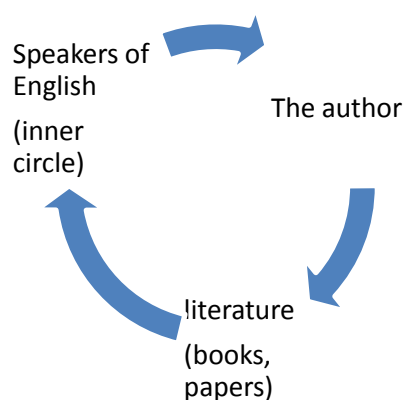


Figure 4.2: Schematic procedure for the pilot phase of this study

This phase of the research is largely characterised by a brainstorming and ‘think aloud’ process. The number of collocational features could be enormous, making it impossible to analyse them in one discrete study. Something relevant and concrete needs to be found, on a manageable scale. The literature acts as a major starting point in the search for target features. They present key differences between Chinese and English in terms of language systems, so they would provide information for type A contrast, as shown in Table 4.4:

Table 4.4: Contrasting procedures

Data	Speaker	Form
Chinese	Chinese	Chinese
English	English and American*	English
CE	Chinese	English

*‘Chinese’ means Mandarin Chinese while ‘English’ is used in a broad sense refers to English “produced in English-speaking countries”.

With reference to information obtained in type A contrast, the next step is to try to see whether there is also similar information in the contrast in B.

The literature consulted that shows type A contrast includes the following:

a. On Chinese grammar

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of

California Press.

Liu, Y. H. (2001). *Shiyong xiandai hanyu yufa [Practical grammar for Chinese language]*. Beijing: Shangwu yinshuguan [Commercial Press].

Lǚ, S. X. (1990). *Zhongguo wenfa yaolue [Key grammars for Chinese language]*. Beijing: Shangwu yinshuguan [Commercial Press].

Wang, L. (1992). *Zhongguo xiandai yufa [Grammar of Chinese language]*. Shanghai: Shanghai shudian [Shanghai Press].

b. On English grammar

Leech, G. N., Deuchar, M., & Hoogenraad, R. (2006). *English grammar for today: A new introduction* (2nd ed.). Basingstoke, England: Palgrave Macmillan.

Leech, G., & Svartvik, J. (2013). *A Communicative Grammar of English*. New York: Routledge.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Quirk, R. (1999). *Longman grammar of spoken and written English*. London: Longman.

c. On the contrast between Chinese and English

Hu, S. Z. (1993). *Yinghan xiuci bijiao yanjiu [Comparative studies in English &*



Chinese rhetoric]. Shanghai: Shanghai waiyu jiaoyu chubanshe

[Shanghai foreign language education press].

Lian, S. N. (2010). *Hanying duibi yanjiu [Contrastive studies of English and*

Chinese]. Beijing: Gaodeng jiaoyu chubanshe [Higher Education

Press].

Pan, W. G. (2010). *Hanying yuyan duibi gailun [Introduction to Chinese*

English contrastive studies]. Beijing: Shangwu yinshuguan

[Commercial Press].

Pan, W. G., & Yang, Z. J. (2008). *Yinghan duibi de lilun yu fangfa yanjiu*

[Theory and practice on English-Chinese contrastive study].

Shanghai: Shanghai waiyu jiaoyu chufanshe [Shanghai foreign

language education press].

Ren, X. L. (1981). *Hanying bijiao yufa [A Chinese-English comparative*

grammar]. Beijing: Zhongguo shehui kexue chubanshe [China social

science press].

Shao, Z. H. (1997). *Yinghan yuyan jiu yu duibi [Researching and contrasting*

English and Chinese language]. Shanghai: Huadong ligong daxue

chubanshe [East China University of Science and Technology Press].

Wang, Z. Y. (2006). *Yinghan yuwen wenti mianmianguan [Aspects on*

English-Chinese contrastive research]. Beijing: Gaodeng jiaoyu

chubanshe [Higher education press].



The process of an early-stage investigation involved reading the literatures, examining relevant samples randomly chosen from the CE data. Notes were made of any features relevant to this thesis, and then speakers from inner circle countries were consulted, to see whether they recognise any of these as potentially salient features based on their contact with Chinese people. This process went back and forth for two rounds until the list was narrowed down to a manageable scope, then the task of “choosing the target to investigate” was accomplished.

The next step consisted of using the software programmes to generate and retrieve patterns of the target words or patterns. From the computer-generated results, the frequencies of occurrences were noted. Then the next task was to identify manually the canonical forms of the target usage of CE, based on two factors:

- a. the relative frequency of occurrences
- b. whether or not these results were due to chance (e.g., all produced by the same persons)

After the frequency information of the chosen words were identified, the next step was to determine whether a similar usage is found in an inner circle variety



(in our case, BrE), with similar frequency of occurrences.

The term “frequency” (Leech, 2011) is associated with three more or less discrete usage patterns, which mean rather different things in corpus linguistics:

- a. raw frequency* (tokens): a count of how many instances of some linguistic phenomenon X occur in some corpus, text or collection of texts;
- b. normalised frequency*: frequency relative to a standard yardstick (e.g., tokens per million words);
- c. ordinal frequency*: a frequency in a list of ranks, in which words are listed in order of frequency after the raw frequency of X is compared with the frequencies of Y, of Z, etc.

Frequency information *b* and *c* were derived from the raw measure *a*. In the following chapter, the type of frequency intended will be specified where appropriate. The scope and target were narrowed down to a certain degree, whereupon detailed studies on collocational features of those specific patterns were generated by the corpus toolkit AntConc (Anthony, 2011). Results are then summed up and will be reported in Chapter 5.

CHAPTER 5

RESULTS: CE COLLOCATIONAL FEATURES

This chapter seeks to explore patterned deviations of CE from one inner circle English variety (BrE) in terms of collocations, with a view to addressing the three research objectives stated in Chapter 2, namely:

1. To provide a descriptive account of the use of CE in terms of collocation;
2. To examine whether the use of these particular collocational features exhibits any distinguishing characteristics marking the different linguistic backgrounds and worldviews of Chinese users of English in general; and
3. To explore the possibility of advancing a theoretical framework for researching CE collocational features.

It is important to note here that collocation can be identified either broadly or narrowly, i.e., either in terms of mere co-occurrence in a numerically defined search horizon (usually 2 to 4 items to the left and 2 to 4 items to the right of the keyword), or in terms of actual pairs of words as modified items. There is a



close relationship between these two ways of defining collocations, since they may be interwoven and partly overlap, this research only considers collocates in the narrow sense for further manual analysis, that is, recurrent word combinations.

5.1 Research focus: selection of items for analysis

The scope of investigation is first justified by outlining the procedural steps of what the targets for study are, as well as the reasons for the choice.

5.1.1 Procedural steps

The identification and description of procedural steps for corpus studies for collocations, i.e., the co-occurrence patterns of specific words, followed the steps as laid out in Table 5.1 (adapted from Sinclair, 2003, pp. xvi-xvii):



Table 5.1: Procedural steps of identification of patterns

Step	Action
1. Initiate	Observe repeated words in right one position (R-1) of the node word; observe repeated words in left one position (L-1) of the node word; identify 'strongest' pattern based on judgment/individual discretion; initiate starting point for 'collocation'.
2. Interpret	Repeated words from the same class/have similar meanings, form a hypothesis.
3. Consolidate	Seek support of the hypothesis established so far by: looking beyond further position (e.g., R-2, L-2); looking at the adjoining words and even some more distant ones; considering words on the other side of the node.
4. Report	Try to exhaust observations of patterns of the node; revise the initiated pattern in the first three steps; write an explicit, testable version for research.
5. Recycle	Go through the same steps as before; look for the strongest pattern remaining on either side; add on patterns found later.
6. Result	Make a final list of hypotheses; link them in a final report on the node.
7. Repeat	Start by applying the report to new data (similar data); go through the same steps; confirm, extend or revise the hypotheses.

After completing the first round of pilot search and consulting references (process described in section 4.4 above), corpus searching and manual work were alternated until results were reached. This chapter describes the findings concerning prevalent preferences of CE users as prelude for delineating their unique features.

5.1.2 Node word, concordance and KWIC

It is important to be clear from the outset about how we target the search items technically. Firstly, suitable target words need to be identified for the search. There is general agreement among corpus linguists that it is easiest to get started by choosing a certain part of speech, one of a group of traditional classifications of words according to their functions in context, including noun, pronoun, verb, adjective, adverb, preposition, conjunction, etc. Accordingly, we will look at potential target items (or node words in corpus linguistics) that belong to a certain grammatical category. Having selected the node word, we can then make a concordance (a list of individual word forms in a given specific context) with the aid of corpus tools, which will then provide us with the keyword in context (KWIC) of the node word. A demonstration of what this may look like in its most basic form is provided below, using AntConc (Anthony, 2011) as the tool, and tiny segment of the CE data.

Take for example the target node word *revolutionary*. This node word was processed to produce a concordance, as shown in Figure 5.1 below. When user-nominated node words are sorted, they appear firstly in the central position, with their co-texts around them, ready to be re-sorted by user-customised search criteria.

Concordance	Concordance Plot	File View	Clusters	Collocates	Word List	Keyword List
KWIC						
<p> wo other facts. In the 1930s, a revolutionary army known as the Red Women Detach- founded on January 1, 1912 by a revolutionary called Sun. By a strange coincidence wn village, but it has become a revolutionary shrine, the site where headquarters University, which used to be a revolutionary venue with a long history where there including ancient cultural and revolutionary monuments, scenic spots, convales- Daiying and Li Hanjun conducted revolutionary activities, has opened to the pub- y unveiled her newest cookbook, Revolutionary Chinese Cookbook: Recipes from Hu- her was imprisoned as a counter-revolutionary in 1969. Jin was sent to the harsh ngkai (1877-1925), a democratic revolutionary, and his wife, He Xiangning. In the ite, it has now become a famous revolutionary tourist attraction. A lot of histo- tures reflecting Sun's glorious revolutionary life. The walls of the hall are d- to March 1947. As an important revolutionary site, it has now become a famous nary historical sites including revolutionary base areas such as Jinggangshan in- most of which are regarding its revolutionary history. In the garden some of the region but also illustrate its revolutionary spirit. With such an abundance of cticed medicine and carried out revolutionary activities in Macao. A full size first golf course built in post-revolutionary China was designed by golfing icon Mountains were the first rural revolutionary base area set up by the late Chai ic government during the Second Revolutionary War; and Anyuan, one of the birth- ding of New China in 1949, some revolutionary historical sites including revolu- ey're eloquent reminders of the revolutionary fervor of the 1950s and 1960s. So eek lines of the lobby with the revolutionary fluid roundness of its furniture. oll. Nancy Zhang looks over the revolutionary new building. The maverick archit- omantic aura, a reminder of the revolutionary passion of those days when people </p>						

Figure 5.1: KWIC analysis as exemplified by the keyword *revolutionary*

5.1.3 Tokens, types, lemmas: definition of terms

Tokens in a corpus refer to the running total of all words, including recurrences, while the number of types refers to the total number of unique words. So the sentence ‘*It is not easy to say it correctly*’ would contain eight word tokens and seven word types (because the word *it* appears twice).

The canonical form of a word is known in corpus linguistics as a lemma. Francis and Kucera (1982) define a lemma as a “set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling” (p. 1).

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

People working with corpus data usually use a kind of automatic annotation that is closely allied to the identification of parts of speech, and which leads to the reduction of the words in a corpus to their respective lexemes, called

lemmatisation. Though not without problems (e.g., *bear*, *bore*, *born* which are verbs, together with *bear* as one kind of animal, would be lemmatised in one lemma by the computer), this process allows the researcher to extract and examine all the variants of a particular lexeme without having to input them all, and to produce frequency and distribution information for the lexeme as a whole. In this study, care was taken to avoid choosing homonyms such as *bear* as target words, and so the target words chosen in this study do not involve such problems.

5.1.4 High frequency nouns in inner circle English varieties

Choosing target word lemmas is a procedure which cannot be performed without first making a number of decisions. In the case of the current research, the decision was made to focus on high frequency nouns, for these reasons:

The pilot study described in Chapter 4 revealed that collocational features seem to emerge around nouns. A second reason is the high density of nouns in English. According to Kennedy (2003), who conducted a corpus analysis of a similar type as the present research, nouns are predominant in English, as shown in Table 5.2:

Table 5.2: Frequencies of word classes in written American and British English

Word class	Total (1 million words in each corpus)	
	AmE (%)	BrE (%)
Nouns	26.80	25.50
Verbs	18.20	17.80
Determiners	14.16	14.20
Prepositions	12.04	12.20
Adjectives	7.07	7.30
Pronouns	6.56	7.10
Conjunctions	5.92	5.50
Adverbs	5.23	5.50
Others	4.02	5.20
%	100.00	100.00

(Source: Kennedy, 2003, p. 60)

The relative distribution of word classes throws further light on the process of choosing targets. It is worth noting that over a quarter of the words of both corpora are nouns. Consequently, nouns make up the highest proportion of words conveying information, which are particularly suitable for serving as the target for the current research.

Are there salient features in CE speakers' use of nouns? The first step to establish a firm basis for addressing this question is to identify high-frequency

nouns in inner circle varieties of English. As is demonstrated by West's *General Service List* (1953), Coxhead's (2000) *Academic Word List* and Nation's (2001) *University Word List*, etc., a vast majority of texts are composed of a relatively small number of very high-frequency words.

All of these works are consistent with the mathematical work of Zipf, known as Zipf's law (1935, 1949), which gave heavily weighted importance to the most frequent words. Zipf's law states that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. So the ranking of a particular word is directly related to its occurrences/frequency. Zipf's law is highly relevant to the selection of research target in this thesis, which is why we choose the high-frequency words as the focus of this study. As Leech (2011) observes, the most frequent 50 word types account for 50% of the word tokens in a corpus, so "frequency can no longer be ignored" (p. 4).

With this insight as the dictum that guides this investigation, I examined the high-frequency words in inner circle varieties. Table 5.3 shows nouns with the highest frequency in British English and American English, as lemmas:

Table 5.3: The 37 Lemmas shared by BNC and COCA (of the top 50)

#	Shared lemma (nouns)	Frequency per million words (BrE)	Frequency per million words (AmE)
1	TIME	1,833	1,699
2	YEAR	1,639	1,709
3	PEOPLE	1,256	1,537
4	WAY	1,108	1,045
5	MAN	1,003	911
6	DAY	940	962
7	THING	776	890
8	CHILD	710	742
9	MR	673	419
10	GOVERNMENT	670	425
11	WORD	653	340
12	LIFE	645	740
13	WOMAN	631	759
14	SYSTEM	619	445
15	CASE	613	446
16	PART	612	462
17	GROUP	607	510
18	NUMBER	606	413
19	WORLD	600	674
20	AREA	585	368
21	COMPANY	579	452
22	PROBLEM	565	484
23	PLACE	534	450
24	HAND	532	501

25	SCHOOL	529	676
26	COUNTRY	486	496
27	POINT	484	394
28	WEEK	476	443
29	STATE	440	605
30	FAMILY	428	541
31	FACT	426	365
32	MONTH	398	362
33	BUSINESS	394	343
34	NIGHT	393	410
35	EYE	392	376
36	HOME	390	379
37	QUESTION	390	427

These frequency figures were adapted from Kennedy (2003, p. 102), who summarises data from the BNC, and from Davies and Gardner (2010), who identify frequencies using COCA, the 450-million word Corpus of Contemporary American English. Both of these corpora are representative enough in terms of size and structure for British English and American English.

Among the top 50 high-frequency nouns, 37 of them are shared, providing some evidence of variation in frequency even between groups of native speakers. While acknowledging that inner circle varieties are good examples to follow for learning English, sticking too rigidly to them may not be a good idea.



For example, the lemma STUDENT occurs 567 times per million words in American English, but not as frequently in British English (222 per million). Likewise, the lemma HOUSE ranks 20th in the BNC list (598 per million), but does not occur as frequently as in COCA (332 per million). The word SERVICE has a frequency of 325 per million words in COCA, a much lower frequency of occurrence compared to the frequency in BNC (549 times per million words). Frequency counts may reveal the relative significance of a target word, but it does not imply that ‘STUDENT’ or ‘HOUSE’ is less important or unimportant for L2 learners, so numbers generated in frequency counts would usually need further investigation before feasible implication and interpretation can be reached.

More examples could be found if careful observations were made. Deviation in frequency of use is common even within inner circle varieties. Whether we like it or not, it seems that frequency information of a lemma allows us to deduce the probability of encountering that lemma – a factor that matters most in communication. For example, the following is an AntConc-generated wordlist of a news story, named *Search ends for 2 passengers believed to have fallen off cruise ship*, randomly chosen from the CNN news website (Retrieved from http://edition.cnn.com/2013/05/10/world/asia/australia-missing-cruise-passengers/index.html?hpt=ias_c2).



Rank	Freq	Word
1	22	the
2	10	and
3	8	of
4	7	a
5	7	ship
6	6	cruise
7	6	in
8	5	carnival
9	5	off
10	5	said

Figure 5.2: A sample wordlist based on one news story

The first column on the left is ‘ranking’, indicating the order of frequency of words. The second column to its right is ‘Freq’, which represents the frequency of each word in the corpus. So interpretations of the first two lines are: the word *the* occurs in that news headline 22 times and the word *and* occurs 10 times. They rank number 1 and number 2 in the outcome list. Words ranked top ten in any wordlist like this one would almost always be those words like *the*, *of*, *a*, *and*, etc., which constitute the key substance in communication. Together with the information that is being conveyed, in this specific case we can see that the content words *ship*, *cruise*, *carnival* make up the theme of the news story. In a word, high-frequency words contribute to our understanding of language use and have been brought to the attention of linguists.

5.1.5 Narrowing down targets

However, there is more to be done with this initial pool of 37 lemmas before the targets for analysis were identified. Two more criteria were added to the choice of words selected for scrutiny. Firstly, the target words should belong to only one word class, e.g., *noun*, in order that any possible ambiguity be minimised:

e.g., SEA ✓
BOOK ×

In the two lemmas, SEA and BOOK, the former fits in the criteria of being in only one word class, while BOOK is not. BOOK could be a noun as well as a verb. Computer programmes could not distinguish a noun from a verb automatically, and manual distinctions are too tedious due to their high frequency of occurrences. In this way, some lemmas were deleted as targets for research, so that results would not be skewed.

Secondly, the selected words should be general words that cover all categories of subjects in CE data with high-frequency ranking, so that the representativeness of their usages is guaranteed.

e.g., CITY ✓
TEMPLE ×



These two lemmas are similar in that they both refer to some places, but they differ in terms of frequency. CITY is predictably higher in frequency compared with TEMPLE, so the collocations of TEMPLE are not representative enough because they only occur once in a while.

After applying these two criteria, twenty-four lemmas in this list were deleted upon closer scrutiny. Firstly, words that can sometimes be used as some other part of speech categories, e.g., PART, GROUP, PLACE, HAND were excluded. The second group of deleted words is related to the notion of ‘time’, which is not of interest in this research. One reason is that they involve more than one grammatical interest to linguists and could be looked at from many different angles. The lemmas YEAR, MONTH, DAY, WEEK, NIGHT were discarded for this reason. Thirdly, the plural form of a lemma such as LIFE has the same form as the verb ‘lives’ in the sentence ‘He lives well’, thus making it potentially ambiguous, and is therefore avoided in this study. Lastly, words like MR. and STATE were not relevant to my research interest and therefore excluded. The reason is that the focus of this research is on the patterns of CE words co-selecting each other in a general way. These words are topic-specific (e.g., in addition to being a homonym involving the verb *state*, as a noun, STATE is used in AmE while the counterpart for this is PROVINCE in CE, these two lemmas are topic-specific in context) and thus do not fit in here.



Because when ‘Mr. Black’ or ‘United States of America’ is addressed, they are not linguistic patterns. The final list of target lemmas for investigation is as follows (Table 5.4):

Table 5.4: Target node words chosen (in descending order as found in BNC)

No.	Lemma	BNC	CEC
1	PEOPLE	1,256	1,543
2	WAY	1,108	1,096
3	MAN	1,003	874
4	THING	776	515
5	CHILD	710	408
6	SYSTEM	619	967
7	AREA	585	673
8	PROBLEM	565	443
9	COUNTRY	486	687
10	FACT	426	405
11	EYE	392	284
12	HOME	390	382
13	QUESTION	390	352

The information of their occurrences in contrast to each other was summarised in Figure 5.3.

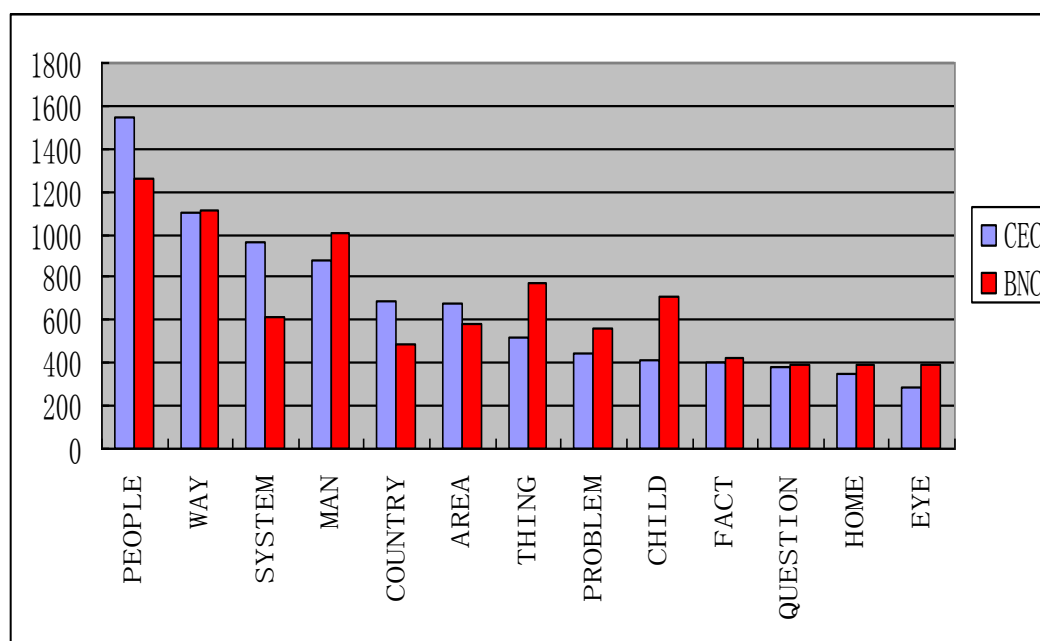


Figure 5.3: Comparative frequency data for target node words (per million words)

These are common words in daily written and spoken use which do not have ambiguous usage in terms of word class. All of them are countable nouns, which mean that they have two word types, singular and plural, facilitating computerised investigation. During the process of researching collocations, there is often doubt as to whether, for example, a word ending in *-ed* or *-ing* is a verb or an adjective derived from a verb, when they stand alone as one word (e.g., *-ing* in *sing*, *singing*). In this thesis, doubts of similar kinds will be dealt with manually by checking their context in the concordances, and sometimes rather arbitrary decisions have to be tolerated. Owing to the need to set a workable limit to the scope of investigation, only high-frequency results were

retained. So the following results are based on, but not limited to, two-word collocations. The steps involved in AntConc search were outlined in Table 5.5.

Table 5.5: AntConc operations

Action	Purpose	Result
a. wordlist	obtain frequency information	target a word
b. KWIC	obtain the concordance for this word	file of this word
c. clean	manually delete irrelevant entries	prepare file
d. save	save output in text file	file of results produced

By alternating between examining the frequency lists and their KWIC data, the word order focus was progressively narrowed down. Our understanding of CE patterns will be further enhanced through follow-up analyses, as the investigation proceeds according to the research framework (see Chapter 4).

In accordance with the literature reviewed in Chapter 2, some hypotheses may be put forward with regard to CE collocational features for a start of the investigation. For example, transference of the linguistic pattern in Chinese (L1 for CE users) (e.g., a modifier-modified order) may be one of the marked features in noun phrase collocations in CE; Overusing genitive case as modifier of nouns may constitute a unique collocational feature in CE because using

genitive case as modifier of nouns has been a dominant rhetorical device in the collocations of nouns; in traditional Chinese culture, views toward philosophy and social values tend to be homogeneous, which may be reflected in collocations in CE – their meaning may be self-evident among Chinese people but foreign to speakers/writers of other varieties of English.

The identification of recurrent patterns distinctive to CE was processed according to the planned procedure and framework described in Section 4.1 and 4.4. Following Sinclair (1991, p. 117), Kennedy (2003, p. 472), Kilgariff (2006, p. 999), Stubbs (2009, p. 127), high-salience collocates should be set within a certain span in the search; CE collocations, in this sense, were identified. They are categorised into two types – overused and unique CE collocations – in the following section.

5.2 Results: CE collocational features

According to Granger (2009, p. 23), the language used by learners of English could be roughly assigned to categories of ‘overuse/underuse/misuse’ (as shown in Figure 5.4). Obviously, the ‘misuse’ could be bracketed off as ‘errors’, which are in need of correction, and which is the target for research in EA. The ‘overuse’ and ‘underuse’, then, are what we need to examine closely in order to



decide whether they should be ‘ignored’ or ‘selected’ for language teaching (Figure 5.4). In what follows, results of overused and underused collocations will be presented.

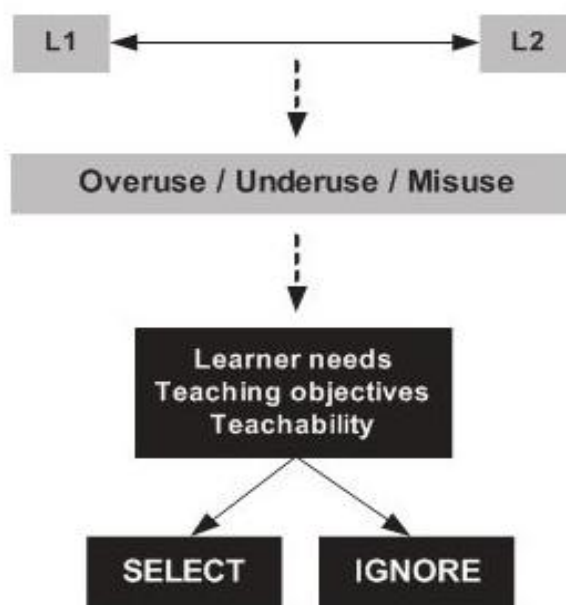


Figure 5.4: From learner language analysis to language teaching (Granger, 2009, p. 23)

5.2.1 Collocations

We shall begin by looking at a profile of the frequencies of occurrence of the overused collocations found in the CEC data. By ‘overused’ I mean while the collocations are both used in CEC and BNC, their numbers of occurrence are very different, suggesting that the collocations in question are preferred, thus constituting evidence of ‘preference’ by CE users.

5.2.1.1 Overused collocations

A list of overused collocations in CE (compared to BNC) has been identified.

Owing to space limitations, we will focus on the top 20 overused collocations in the CEC (in descending order of CEC/BNC ratio) (see Table 5.6).

Table 5.6: 20 overused collocations in CE

Overused collocation	No. of tokens per million in CEC	No. of tokens per million in BNC	CEC/BNC ratio	Ranking
minority AREA	4.67	0.01	467	1
poverty/earthquake/ disaster-stricken AREA	6.11	0.02	305.5	2
superior MAN	3	0.02	150	3
western PEOPLE	2.6	0.02	130	4
bluest EYE	3.03	0.03	90.8	5
responsibility SYSTEM	4.67	0.06	77.8	6
coastal AREA	8.56	0.19	45.1	7
poverty-stricken PEOPLE	1.25	0.03	41.7	8
foreign PEOPLE	1.25	0.02	41.7	9
autonomous AREA	3	0.08	37.5	10



agency MAN	0.58	0.02	29	11
mortal EYE	0.31	0.01	27.5	12
island COUNTRY	1.03	0.04	25.8	13
earthly THING	0.6	0.03	20	14
genuine QUESTION	0.6	0.03	20	15
comprehensive WAY	2.41	0.13	18.5	16
scientific WAY	1.88	0.14	13.4	17
existing PROBLEM	2.28	0.17	13.4	18
common PEOPLE	15.27	1.18	13	19
concrete FACT	0.42	0.04	11	20

In all of these cases, there is no question about the acceptability or intelligibility of the collocations, as evidenced in their occurrence in the BNC. However, CE users tend to prefer using a variety of different collocations. Non-conformity of usage here does not occur in the linguistic form of the target words (found in both corpora) but in their frequency (markedly different frequencies of use). The more frequent a collocation is, the more likely it is that people from other cultures would encounter it.

The phenomenon of overuse of some collocations serves as evidence that CE users tend to have a preferred way of expressing meanings, which fall into three



broadly defined groups, namely, collocations related to:

- a) Sociopolitical ideology (Table 5.7);
- b) Culturally conditioned stance (Table 5.8);
- c) Linguistic structure (Table 5.9).

Table 5.7: CE collocations related to sociopolitical ideology in China

CE (and variants)	Chinese	Pinyin
<i>minority AREA</i> <i>ethnic minority AREA</i>	少數民族地區	<i>shǎoshù mínzú dìqū</i>
<i>responsibility SYSTEM</i> <i>land contract responsibility SYSTEM</i> <i>household contract responsibility SYSTEM</i> <i>household responsibility SYSTEM</i> <i>household-based responsibility SYSTEM</i>	承包責任制	<i>chéngbāo zérènzhi</i>
<i>autonomous AREA</i> <i>Autonomous Region Government AREA</i>	自治區	<i>zìzhìqū</i>
<i>comprehensive WAY</i>	全面	<i>quánmiàn</i>
<i>common PEOPLE</i>	人民 老百姓	<i>rénmín</i> <i>lǎobǎixìng</i>

The first group concerns examples like the slogan *responsibility SYSTEM*. This refers to a kind of policy aiming to motivate peasants to increase their agricultural productivity. In other words, it is the name of a policy, like the well-known *reform and open policy* (also called *the open-door policy*), or the *one-child policy*, *three represents*, and the *one country two systems policy*. *Responsibility SYSTEM* has many variants, as *land contract responsibility system*, *household contract responsibility system*, *household responsibility system* and *household-based responsibility system*. These are Chinese-specific policies, so it seems obvious why these English terms are needed.

Every culture has some distinct practices to be expressed. They are either related to concrete objects (e.g., *hutong* means ‘street’) or abstract entities (e.g., *guanxi*, with a similar meaning to *backdoor practice*, Y. G. Gu, 2002, pp. 273-283). Expressions like *responsibility system* and *one-child policy* are unique in the way that they are created by and specific to Chinese people in Mainland China and used (e.g., in media reports) among speakers/writers across the three Kachruvian circles of English. Over time, they may eventually enter the English lexicon, as shown in other examples like *long time no see*, *dimsum*, and *kungfu* (cf. Xu, 2010; Cummings & Wolf, 2011). These collocations are strongly associated with the sociopolitical settings of China especially by contrast with their British counterparts.



It is worth further displaying examples of the latter two collocations *comprehensive WAY* and *common PEOPLE*, to justify why they were put into this category. Examples for *comprehensive WAY* were taken from the *Report on the Work of the Government*, submitted to the National People's Congress (http://www.gov.cn/test/2006-02/16/content_200719.htm). An official Chinese version is obtainable, which helps minimise the translator's influence.

1. Since the return of Hong Kong and Macao to the embrace of the motherland, the principle of “one country, two systems” and the basic laws of the two regions have been implemented in a **comprehensive way**.

Chinese version: 香港、澳门回归祖国以来, “一国两制”方针和基本法得到**全面**贯彻执行。

2. The development of science, technology and education was accelerated, and other social undertakings progressed in a **comprehensive way**.

Chinese version: 部门所属应用型科研院所企业化改革基本完成, 其他科研院所体制改革**全面**展开。

Table 5.7 also shows that *common PEOPLE* is used more often than its synonym, *ordinary PEOPLE*. A contrast between CE and BNC and Chinese is thus made and shown in Figure 5.5:



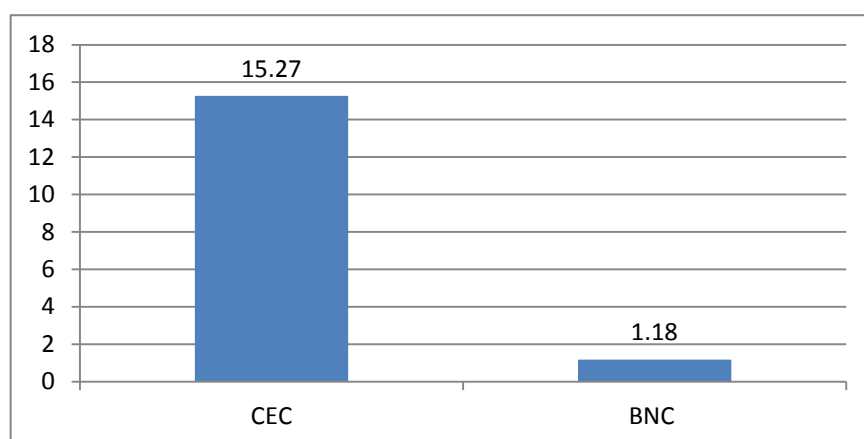


Figure 5.5: Contrasting the frequencies of *common PEOPLE* in CEC and BNC (per million words)

While Figure 5.5 is a CE-BrE contrast, a further contrast of 人民 (*rénmín*) and 人們 (*rénmēn*), both meaning ‘people’ in Chinese was made, using the Chinese version of *Report on the Work of the Government* (2000-2013) as data. The result is shown in Figure 5.6.

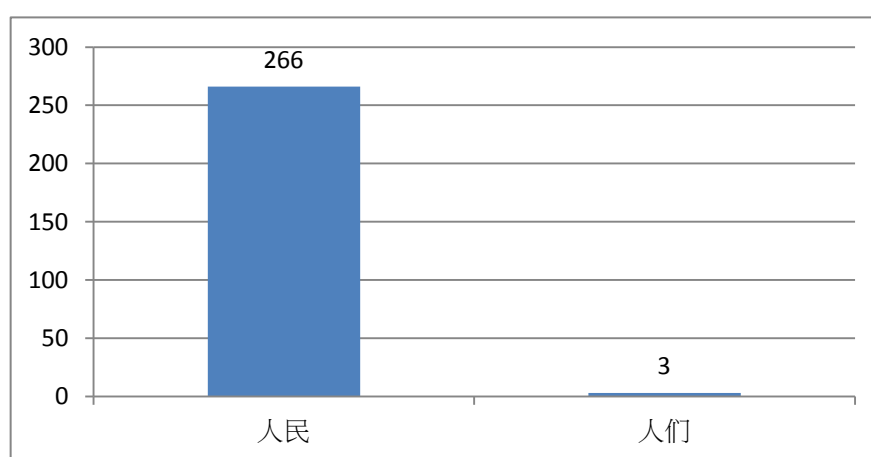


Figure 5.6: Contrasting the frequencies of 人民 (*rénmín*) and 人們 (*rénmēn*) (absolute figures)

In addition, the currency used in Chinese is named 人民幣 *Renminbi* (RMB), which is well known (but not *人們幣) and which carries an important ideologically loaded meaning in the name of the currency itself. It thus seems reasonable to put the overused pattern *common PEOPLE* under the sociopolitical ideology category. Then we may come to the second observation under the overused collocations in CE:

Table 5.8: Collocations related to culturally conditioned stance

CE (and variants)	Chinese	Pinyin
<i>superior MAN</i>	君子	<i>jūnzǐ</i>
<i>scientific WAY</i>	科學的方法	<i>kēxué de fāngfǎ</i>
<i>existing PROBLEM</i>	存在的問題	<i>cúnzài de wèntí</i>

The second category pertains to the culturally conditioned stance or belief of CE users. Consider the collocational pattern *superior MAN*, the root of which dates back to the well-known Confucius sayings (*Analects of Confucius*), based on Taylor and Legge (Translator, 2011), three of which are relevant here:

1. The **superior man** bends his attention to what is radical. That being established, all practical courses naturally grow up. (Chapter 2)

Chinese version: 君子務本，本立而道生。(第二章)



2. Tsze-kung asked what constituted the **superior man**. The Master said, ‘He acts before he speaks, and afterwards speaks according to his actions’.

(Chapter 13)

Chinese version: 子貢問君子，子曰：先行其言、而後從之。（第十三章）

3. The Master said, ‘The way of the **superior man** is threefold, but I am not equal to it. Virtuous, he is free from anxieties; wise, he is free from perplexities; bold, he is free from fear’. (Chapter 30)

Chinese version: 子曰：‘君子道者三，我無能焉：仁者不憂，知者不惑，勇者不懼’。（第三十章）

As shown in these excerpts, a ‘*superior MAN*’ (*jūnzǐ*) is an idealised Chinese person that is held in high regard in traditional Chinese culture for centuries. There is no equivalent of English for *superior MAN*, as stated and described in the *Analects*, which has been passed on from one generation to the next and is pervasive in Chinese people’s thinking, belief, and behaviour as well to some extent. Thus collocations like ‘*superior MAN*’ are culturally salient to CE users, which is a reflection of thinking style of Chinese people. Other collocations that users from other cultures may find redundant are *scientific WAY* and *existing PROBLEM*, among many others.



In the case of the overused collocation *scientific WAY*, we can see a shared understanding of the world, which is the basis for shared knowledge of lexical and collocational preferences in communication among CE users.

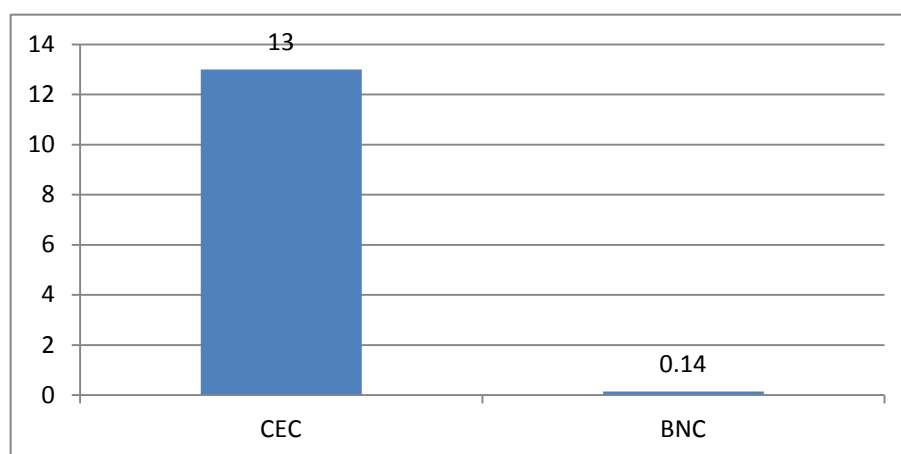


Figure 5.7: Contrasting the frequency of *scientific WAY* in CEC and BNC (per million words)

Further investigation shows that the overused feature of this particular collocation comes from the overuse of the word *scientific*. This word is homogenously positive in Chinese culture, especially after the two *Opium Wars* (1839-42; 1856-60) which were commonly blamed on Imperial China's backwardness in science and technologies. In modern China, being *scientific* is always advocated, as can be seen from the contrast of its frequency of occurrences in the two corpora in Figure 5.8, and KWIC of one CE collocation *a scientific outlook on development* in Figure 5.9.

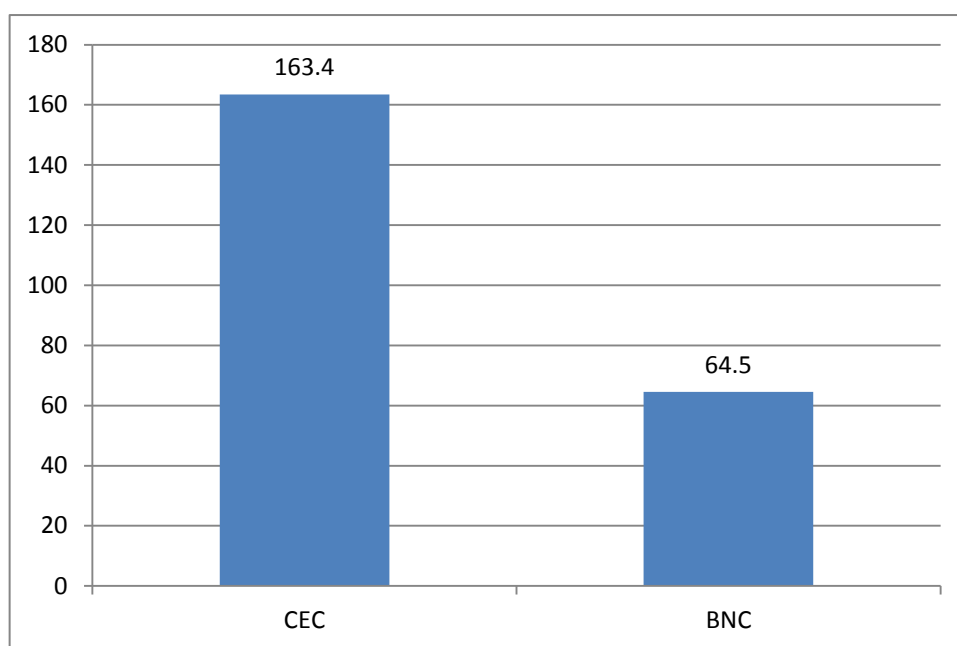


Figure 5.8: Contrasting the frequencies of *scientific* in CEC and BNC (per million words)

Scientific occurs a lot more frequently in CEC than BNC. That ‘scientific’ is a kind of yardstick for advancement and progress, both being arguably indispensable for social prosperity and personal happiness, is truly a deep-rooted belief. The choice of this word as the modifier is thus preferred. Or, in corpus linguistic terms, *scientific* has higher collocationality (Kilgarriff, 2006) in CE than in BrE.

actual circumstances? A scientific outlook on development is a guiding principle to establish and implement a scientific outlook on development. This outlook is a new development; improving the army; take a scientific outlook on development as an important guiding principle; social development by a scientific outlook on development; to strengthen and improve the committee, fully implement a scientific outlook on development, continue to accelerate (draft) fully reflects a scientific outlook on development and embodies the importance. We must be guided by a scientific outlook on development in solving current outstanding problems. Society We must adopt a scientific outlook on development and implement the strategy of implementing a scientific outlook on development and maintaining a correct philosophy of implementing a scientific outlook on development and building a harmonious society. It is a way of implementing a scientific outlook on development over the past two years. We adopted a pragmatic approach, a scientific outlook on development and a correct attitude towards the efforts to implement a scientific outlook on development and to tighten and improve in all respects. We adopted a scientific outlook on development to guide China's overall development." We will maintain a scientific outlook on development and a correct attitude towards

Figure 5.9: KWIC of *a scientific outlook on development* in CEC

The collocation *a scientific outlook on development* was put forward by the former General Secretary of the Communist Party of China, Hu Jintao, at the Third Plenary Session of the 16th CPC National Congress on 28 July 2003, and it has been used to guide the further development of China ever since. We may see from this innovative collocation, together with other discussion above, that *scientific* is a preferred word in the CE users' lexicon, resulting in an observable and to some extent predictable high-frequency occurrence in CE collocations.

The third category concerns CE users' collocational preferences which have

been transferred from the patterns of expressing ideas in Chinese. To illustrate, six examples were chosen, which represent two types of linguistic structure: hyphenated compound adjectives and distinct usage of two common adjectives (i.e., *western* and *foreign*).

Table 5.9: Linguistic structure feature

CE	Chinese	Pinyin
<i>poverty-stricken AREA</i>	貧窮地區	<i>pínqióng dìqū</i>
<i>earthquake-stricken AREA</i>	地震災區	<i>dìzhèn zāiqū</i>
<i>disaster-stricken AREA</i>	災區	<i>zāiqū</i>
<i>poverty-stricken PEOPLE</i>	貧困人口	<i>pínkùn rénkǒu</i>
<i>western PEOPLE</i>	西方人	<i>xīfāngrén</i>
<i>foreign PEOPLE</i>	外國人	<i>wàiguórén</i>

One rhetorical device for which CE users have a preference is their predilection for hyphenated compound adjectives, as shown in examples like *poverty-stricken* and *earthquake-stricken* (cf. Table 5. 6). More innovative and creative examples following this principle include *sadness-stricken*, *tsunami-stricken*, *hunger-stricken*, *epidemic-stricken* and *flu-stricken*. They were searched in the BNC, but no tokens of these collocations were found. On the basis of the hypothesis that use of the hyphen to create compound adjectives

is an overt feature of CE, further searches were performed and more evidence has been found. One example is the hyphenated compound adjectives *all-round*, whose ratio per million words in both corpora is also highly marked (Figure 5.10):

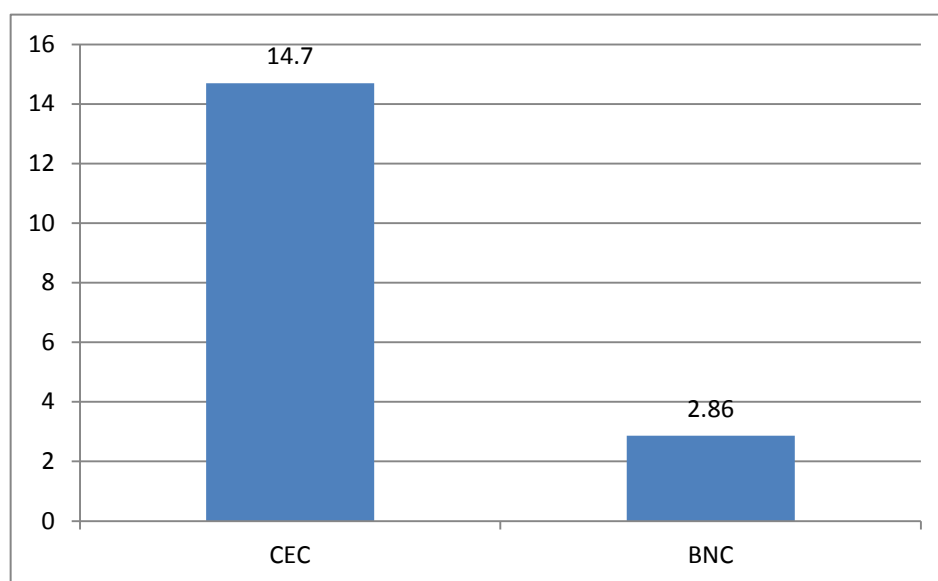


Figure 5.10: Contrasting the frequencies of *all-round* in CEC and BNC (per million words)

Like the word *scientific* discussed above, hyphenated compound adjectives as a rhetorical device seem to have higher collocationality in CE than in BrE. The second type of examples that can serve as linguistic preference is *western PEOPLE* (meaning people from Western countries, i.e., foreigners). The contrasting ratio of its frequency of occurrence in both corpora is just as revealing (Figure 5.11):

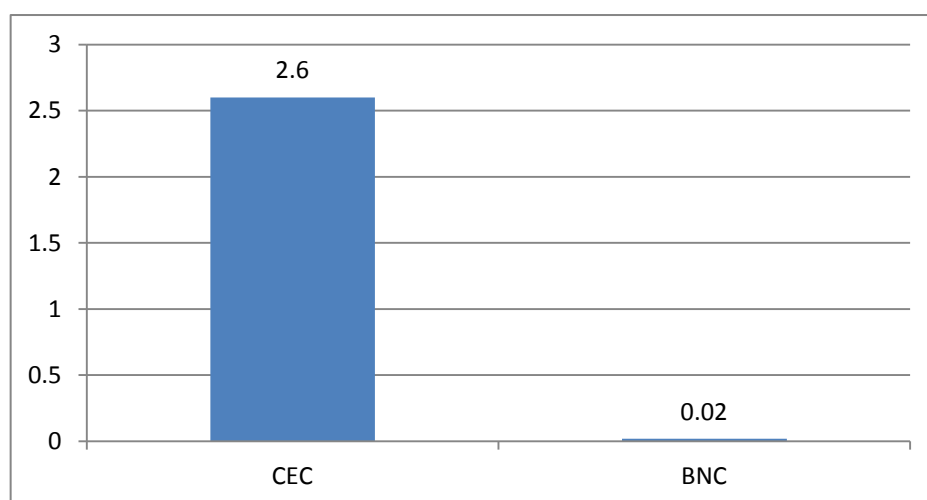


Figure 5.11: Contrasting the frequencies of *western PEOPLE* in CEC and BNC (per million words)

Western people in BNC:

- a) This latter attitude is not unique to modern times, or to **western people**.
- b) He realized gloomily that individually, **Western people** suffer from debt just as Third World people do.

Western people is an acceptable way of referring to non-Chinese people (in Chinese contexts in general). It seems reasonable to argue, as these two overused collocations suggest, that choice of collocations arises naturally not so much because their users have difficulty in finding the right words. Rather, it should be pointed out that CE users are able to use the alternative *people from the west* as well. However, the data shows that they clearly have a predilection for *western people*. Likewise, *foreign people* belongs to a similar category.

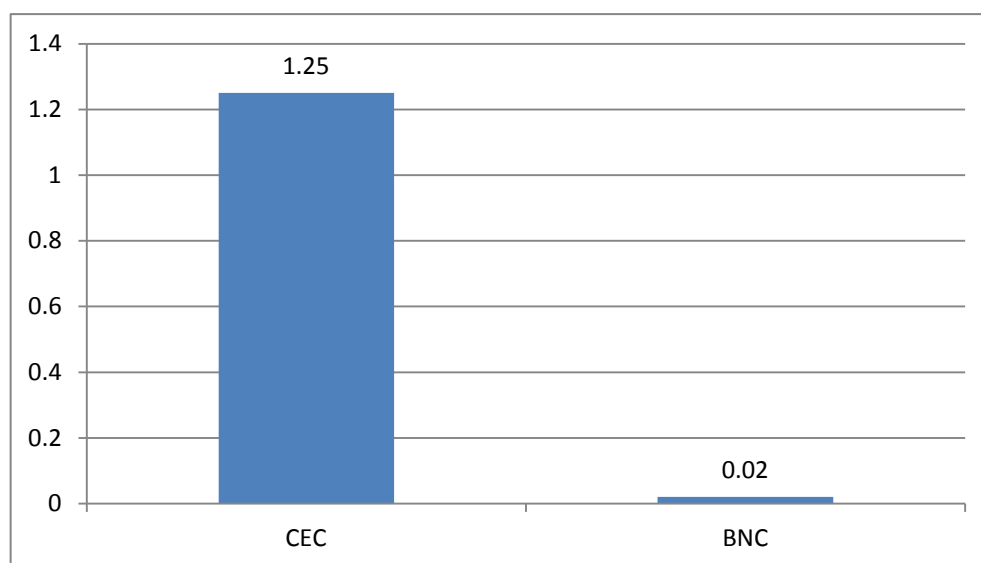


Figure 5.12: Contrasting the frequencies of *foreign people* in CEC and BNC (per million words)

Foreign people in BNC:

1. In fact, as the global expansion of Indian and Chinese restaurants suggests, xenophobia is directed against **foreign people**, not foreign cultural imports.
2. It will describe and analyse how pupils perceive **foreign people**, in particular the French, and which environmental factors, including foreign language teaching, influence their perceptions.

To demonstrate the collocationality of the preferred modifier (or device), nouns that are modified by *western* and *foreign* (span: R-1) are contrasted below (Tables 5.10 and 5.11).

Table 5.10: Top ten nouns modified by *western* in CEC (per million words)

Noun	tokens
REGION	21
COUNTRY	8.3
CULTURE	6.6
PART	3.1
PEOPLE	2.6
CHINA	1.8
WORLD	1.8
AREA	1.5
STYLE	1.5
CIVILISATION	1.3

Table 5.11: Top ten nouns modified by *western* in BNC (per million words)

Noun	tokens
EUROPE	11
WORLD	3.5
COUNTRY	2.4
ISLE	2
SOCIETY	2
AUSTRALIA	1.4
RAILWAY	1.3
CULTURE	1.3
GOVERNMENT	1.1
FRONT	1.1

Through this comparison, the distinct usage of five CE collocations with *western* (as highlighted in Table 5.9) was attained (Table 5.12).

Table 5.12: CE collocations with *western*

CE	Chinese	Pinyin
<i>western REGION, AREA</i>	西部地區	<i>xībù dìqū</i>
<i>western PEOPLE</i>	西方人	<i>xīfāngrén</i>
<i>western STYLE</i>	外國風格	<i>wàiguó fēnggé</i>
<i>western CIVILISATION</i>	西方文明	<i>xīfāng wénmíng</i>

It may be that these collocations, with *western* as the modifier, sound more natural and comfortable both for speakers and listeners engaged in ELF communication. There are other alternative ways to express these terms, for example, *people from western countries*, *people from foreign countries*, *foreigners*, *exotic*, etc. However, as shown in the frequency-based information above, the version using *western* as modifier is preferred, thus constituting evidence of CE users' preference.

In Table 5.9, there is another overused collocation whose meaning is close to *western PEOPLE*, i.e., *foreign PEOPLE*. The same process of comparison and contrast were made and presented in Table 5.13—Table 5.15.

Table 5.13: Top ten nouns modified by *foreign* in CEC (per million words).

Noun	tokens
LANGUAGE	37
EXCHANGE	27
TRADE	13
INVESTMENT	11
COUNTRY	7.2
CAPITAL	6.9
CURRENCY	5.7
INVESTOR	5.5
POLICY	5.4
COMPANY	3.9
AFFAIR	2.4

Table 5.14: Top ten nouns modified by *foreign* in BNC (per million words)

Noun	tokens
MINISTER	21
AFFAIR	12
POLICY	10
OFFICE	9.4
EXCHANGE	7.6
SECRETARY	6.3
LANGUAGE	5.6
CURRENCY	4.7
MINISTRY	4.5
INVESTMENT	3.5

The comparison generated some distinct usages, as shown in Table 5.15.

Table 5.15: CE collocations with *foreign*

CE	Chinese	Pinyin
<i>foreign TRADE</i>	外貿	wàimào
<i>foreign CAPITAL</i>	外資	wàizī
<i>foreign INVESTOR</i>	外國投資者	wàiguó tóuzīzhě
<i>foreign COMPANY</i>	外企	wàiqǐ

As can be seen, the terms are closely related to the field of economics (e.g., *wàiqǐ*), things that did not originally exist in China but came into being as the society gradually developed. The need to refer to them is by virtue of society development, while the way to refer to it in English is by linguistic preference of CE users nationwide.

In sum, some overused collocations distinctive to CE were found. In general, linguists are accustomed to seeing deviation from English usages in CE as inappropriate English. The three types of CE collocations proposed in Section 5.2.1.1 are intended to demonstrate that researching collocations is one promising direction toward legitimating CE as an expanding circle variety, reflecting its speakers' or writers' lexical and collocational preferences when making meaning.

5.2.1.2 Innovative collocations

Besides overused collocations, a number of unique collocations occur which are only used by CE speakers; they are referred to as innovative collocations in CE. By using the word ‘innovative’, the peculiarity of these collocations is emphasized. One very important argument here towards the innovative collocations is that, the absence of any tokens in inner variety data does not imply that the type (word or phrase) in question is unacceptable or problematic, let alone erroneous.

Given that a ‘complete’ result is not possible, a number of instructive cases will be exemplified below, which fall into two categories:

- a) Hyphenated compound adjectives as modifier and
- b) Collocations reflecting Chinese worldview

- a) Hyphenated compound adjectives as modifier

By hyphenated compound adjective, I mean adjective with two words joined by a hyphen in the L-1 position before the node word. *African-American PEOPLE* and *peace-loving PEOPLE* are examples of these types of innovations (Table 5.16).



Table 5.16: Hyphenated compound adjectives as modifiers in CE

CE	Chinese	Pinyin
<i>African-American PEOPLE</i>	黑人	<i>hēirén</i>
<i>peace-loving PEOPLE</i>	愛和平的人	<i>àihéping de rén</i>
<i>laid-off WORKER</i>	下崗職工	<i>xiàgǎng zhígōng</i>
<i>all-round WAY</i>	全面方法	<i>quánmiàn fāngfǎ</i>

African-American PEOPLE

Although the collocation *African-American PEOPLE* is not found in BNC, it may be counter-intuitive to think that it is inadmissible in BrE. The possibility of an accidental gap in BNC cannot be ruled out. It may happen that this collocation is used by people all over the world, including people in UK, but the data does not include such an example. In order to ascertain (or reject) that *African-American PEOPLE* is innovative in CE, more corpus search was carried out. First, though *African-American PEOPLE* is not found in BNC, the word *African-American* occurs 40 times, indicating that this word is used, though not used in this particular collocation pattern. Then, we searched a corpus of American English (AmE), the 450-million Corpus of Contemporary American English (COCA, <http://corpus.byu.edu/coca/>), in order to get frequency of use of this collocation in AmE. It was found that the word *African-American* occurs 10,359 times in COCA, indicating that this

hyphenated compound adjective is quite a general word in use. However, the collocation *African-American PEOPLE* only occurs 87 times. This is an indication that it is used but the ratio is not so frequent (0.8% in 10,359 tokens). One other way to refer to *African-American PEOPLE*, namely *African-American MAN*, was searched, too, to serve as additional reference for contrast (Table 5.17) (absolute figure).

Table 5.17: Contrasting collocation *African-American PEOPLE* in AmE and CE

Collocation	Tokens: COCA	Tokens: CEC
<i>African-American PEOPLE</i>	87	30
<i>African-American MAN</i>	359	4

These frequencies suggest that *African-American PEOPLE* is innovative in CE compared to BrE, and low in frequency when compared to AmE. One possible reason for this is English teaching in China, as discussed in Section 2.4. *African-American PEOPLE* is taught as one collocation in English textbooks while *African-American MAN* is not. Inputs that are presented in the textbooks and are fostered by the ‘non-native’ teacher of English throughout the years of learning in CE users’ lives, may subconsciously influence the English they use over time.

Peace-loving PEOPLE

A similar analysis applies to the collocation *peace-loving PEOPLE*. In BNC, examples of *peace-loving MAN/NATION/COUNTRY/CITIZEN* were found, but not *peace-loving PEOPLE*. Though significance of this finding is open to interpretation, the descriptive results are somehow indicative of a trend.

The latter two collocations in Table 5.16—*laid-off WORKER* and *all-round WAY*—are indeed Chinese-specific, and needs further elaboration with examples of sentences where they occur.

Below are some examples of *laid-off WORKER*, *all-round WAY* in *Report on the Work of the Government* (2000-2013):

Laid-off WORKER

1. Our targets for 2004 are to create 9 million jobs for urban residents and get 5 million **laid-off** workers reemployed.

Chinese version: 今年的预期目标是, 新增城镇就业 9 0 0 万人, **下岗失业人员**再就业 5 0 0 万人。

2. These centers should pay basic living expenses to **laid-off** workers on schedule and in full and pay the premiums for the social welfare insurance of these workers, including old-age, unemployment and medical insurance.

Chinese version:企业再就业服务中心要向下岗职工按时足额发放基本生活费，并代他们缴纳养老、失业、医疗等社会保险金。

3. The problem of finding other employment opportunities for workers laid off from SOEs was basically solved, and the work of incorporating basic cost of living allowances for laid-off workers into the unemployment insurance system was completed.

Chinese version:基本解决国有企业下岗职工再就业的问题，完成下岗职工基本生活保障向失业保险的并轨。

4. We created jobs through a variety of channels, stepped up vocational and technical training, provided support and assistance to zero-employment households and people who have difficulty finding employment, and helped 5.05 million laid-off workers find new jobs.

Chinese version:多渠道开发就业岗位，加强职业技能培训工作，开展对“零就业家庭”和就业困难人员的帮扶，505 万下岗失业人员实现再就业。

The word *laid-off* is highly Chinese-specific because it describes the status of being unemployed. Workers or employees of the former state-owned factories/enterprises are called *laid-off workers*. Owing to the ‘Centrally Planned Economy’ policy in the 1970s to 1980s, state-owned enterprises in China usually had life-long contracts with all employees, called *zhígōng* (職工),



which is similar to *workers* but different in meaning. The status of being a *zhígōng* indicates that, theoretically, you would get your salary for your whole life, even after retirement, and the salary has nothing to do with how financially successful the enterprises were running, because it is the government that paid the salary. However, following the national open-door policy formally declared in 1978, many state-owned enterprises had to be closed since the mid-1990s due to poor management and unproductive business operations. Thus the groups of *zhígōng* had become *laid-off workers* (*xiàgǎng zhígōng* 下崗職工), as termed by the government. The re-employment of these groups of people is monitored and guided by *the Ministry of Human Resources and Social Security of People's Republic of China*, whose website (<http://www.mohrss.gov.cn/>) reported that in the year 2000, the number of *laid-off workers* had reached a peak of 6.57 million. Against this socioeconomic background, the innovative pattern *laid-off WORKER* merely reflects an important social change in Mainland Chinese society.

All-round WAY

1. Science, technology, education and other social undertakings developed in an **all-round way**.

Chinese version: 科技教育和社会事业**全面**发展。



2. The quality of the teaching staff should be improved in an **all-round way**.

Chinese version: 加强教师队伍建设, **全面**提高教师素质。

3. 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**.

Chinese version: 认真贯彻党的教育方针, 加强德育工作, 推进素质教育, 促进学生**全面**发展。

4. We need to develop all types of human resources in a coordinated and **all-round way**, with the focus on producing innovative scientists and engineers, and experts and professionals in key areas of economic and social development, and energetically attract high-caliber personnel from overseas.

Chinese version: 要**全面**统筹推进各类人才队伍建设, 突出培养创新型科技人才、经济社会发展重点领域专门人才和高技能人才, 积极引进海外高层次人才。

All-round, unlike *laid-off*, is by no means a term denoting Chinese-specific reality. Thus, the 183 tokens of the pattern *in an all-round way* in CEC proved the salience of CE users' preference in a way. The hyphenated compound adjective *all-round* will be further discussed below.



Based on Table 5.16 and its follow-up discussions, a fuller list of innovative collocations using hyphenated compound adjectives as modifiers was reached, as presented in Table 5.18.

Table 5.18: Innovative collocations in CE (frequency: in descending order)

#	Innovative collocations	No. of tokens
1	all-round WAY	183
2	turn-taking SYSTEM	38
3	eco-economic SYSTEM	34
4	African-American PEOPLE	30
5	bare-armed MAN	17
6	peace-loving PEOPLE	15
7	long-coated MAN	8
8	ethnic-minority AREA	7
9	open-minded MAN	6
10	laid-off PEOPLE	5
11	step-by-step WAY	5
12	down-to-earth WAY	5
13	anti-pollution/ anti-corruption/ anti-smuggling SYSTEM	5
14	fund-raising SYSTEM	5
15	multi-nationality COUNTRY	5
16	minority-inhabited AREA	5
17	cotton-producing AREA	5
18	above-mentioned SYSTEM	4
19	community-based SYSTEM	4
20	fast-growing COUNTRY	3



Further investigation of these collocations has revealed that the word *all-round* mentioned in Table 5.18 emerged as one important and active modifier for the nouns under investigation, and differences in usage can easily be detected by checking what they collocate with (Table 5.19, sorted by frequency).

Table 5.19: Contrasting the collocations of *all-round* in CEC and BNC

(in decreasing order of number of tokens, respectively)

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=R-1)

Six of the eight CEC collocates (No. 1, 3, 4, 5, 7, 8, as highlighted) are unique in CE and not found in the whole BNC. Examples involving these collocations are presented and discussed below (*all-round WAY* has been discussed above, thus not included).

All-round COOPERATION:

1. Then he expressed his wishes to launch **all-round cooperation** with Tongji University.

Chinese version:他随后表达了希望与同济大学**全面合作**的愿望。

All-round PROGRESS

- 2 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.

Chinese version:要着力深化改革、提高对外开放水平，着力改善民生促进社会和谐，**全面推进**社会主义经济建设、政治建设、文化建设、社会建设以及生态文明建设。

All-round MANNER

- 3 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.

Chinese version:让我们扎实工作，为把改革开放和社会主义现代化事业**全面**推向新世纪而奋斗！



All-round PARTNERSHIP

- 4 China and Belgium decided on Monday to upgrade their relations to an **all-round partnership** of friendship and cooperation.

Chinese version: 中国与比利时决定，将两国关系提升为**全方位**友好合作**伙伴**关系。

All-round CONSIDERATION

- 5 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.

Chinese version: 环境影响评价必须客观、公开、公正，**综合考虑**规划或者建设项目实施后对各种环境因素及其所构成的生态系统可能造成的影响，为决策提供科学依据。

Since meanings in these examples are clearly made, labelling the very frequently used phrases *all-round WAY* or *all-round COOPERATION* etc., as problematic or bad English is groundless. These collocations cannot be dismissed as errors, too. Because ‘not being used by inner circle varieties’ is a criterion for making judgment in field of EA. Rather than rejecting EA wholesale, our main argument here is that characterising every unique

collocation as an error is totally ungrounded and would be an overly simplistic approach to analysing CE. Two collocations are used by both BNC and CEC in Table 5.19, i.e., *all-round DEVELOPMENT* and *all-round IMPROVEMENT*, which indicate acceptability of usage in British English. Yet we found that the same collocation may be used in a quite different way in the CEC and BNC. Five examples of the collocation *all-round DEVELOPMENT* in each corpus are presented below for contrast:

Examples from CEC (5 out of 49):

1. 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.
2. Zhu reiterated that China will continue its present family planning policy to tackle population problems in China and create a positive environment for **all-round development** in the next century.
3. The good-neighborly friendship and cooperation between China and its surrounding countries have enjoyed **all-round development**.
4. 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.

5. We should continue to increase investment in agriculture, promote **all-round development** of agriculture and the rural economy and work to increase farmers' incomes.

Examples from BNC (5 out of 6):

1. Ensuring the loyalty of the population entails **all-round development**.
2. 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'.
3. Teachers cannot expect to foster a child's **all-round development** if the child's roots are ignored.
4. However, the term general can be conceived of in a third way, in terms of the **all-round development** of the individual.
5. 'We are looking at the **all-round development** of the child, not just narrow academic achievement,' added Ms Robinson.

We may see that though this collocation is used, it is used in different ways or topics. In CEC, the *all-round DEVELOPMENT* is usually about economy, agriculture, or society, while in BNC, the *development* is about people. In other

words, the subjects that *develop* are different.

In sum, as evidenced in the findings and examples illustrated in this section, using innovative hyphenated compound adjectives appears to be one high-frequency collocational feature preferred by CE users. Efforts (e.g., by teaching) to prevent people from using them (and other usages that CE users feel most comfortable with), especially subconsciously, would be in vain. In any contexts of social interaction, people tend to use the language resources that derive from their understanding of their own lifeworld naturally.

b) Collocations reflecting Chinese worldview

By collocations reflecting Chinese worldview, I mean the CE collocations in question are constructed based on its Chinese source or equivalent, hence reflecting its semantic content, typically morpheme-for-morpheme structurally. I will take one of the unique collocations of this kind, *town PEOPLE*, as an example. This collocation is not found in BNC, yet it occurs 34 times in CEC (but *townspeople* are used in BNC). We have found *city PEOPLE* (13 tokens) and *country PEOPLE* (60 tokens, country means countryside) in BrE. To put these existing collocations in linguistic structure, they follow a *collective place (city, country) + people* format. It happens that when CE users want to express



some place not as big as a *city* (*chéngshì* 城市), yet not as small as a *country* (*xiāng* 鄉村), they would eventually use the *collective place + people* format to construct a collocation, which is how they arrive at collocation *town PEOPLE*. While the structure *collective place + people* is used, *town PEOPLE* should be accepted, too. It is one example of a unique usage that speakers of inner circle Englishes may need to accept although they do not actually use it themselves. Given that *city PEOPLE* and *country PEOPLE* are well-attested usages, there is no reason why the collocation *town PEOPLE* should be dismissed as an error and assigned to the list of bad English as target for correction or elimination. Nor is it realistic to stick too rigidly to a so-called ‘perfect’ model of English usage as the goal of interpersonal communication, especially in intercultural settings.

The high-frequency finding *plain MAN and WOMAN* is similarly debatable in its own way. *Plain MAN* is used as in the example from the BNC, showing that this collocation is acceptable:

1. TV criticism, initially a job for the **plain man**, became an important part of quality paper arts pages and attracted star journalists such as Clive James.

However, Chinese users of English use it in innovative and creative ways as shown below (Figure 5.13).



ty and good of the nation, but every plain man and woman, John Smith, editor of the "Patriot
men and women, _even the good honest plain men and women, such men as Professor Dickinson, ins
ump, but it is really the good honest plain men and women, such men as Professor Dickison himse
and Moses Lump belong to the party of plain men and women; and, secondly, because the plain men
not to the mob, not to the crowd of plain men and women, but to their King and their Conscier
of the mob, the panic of the crowd of plain men and women by giving them power. The tragic hope
of the mob, the panic of the crowd of plain men and women; in fact, as I said because the panic
omen, from the panic of the crowd of plain men and women, in fact, from the panic of the mob i
ie mob, fear and worship the crowd of plain men and women in their country. The Russian Emperor
m the mob, the panic of the crowd of plain men and women which makes them helpless. In fact, r
ut down the clamour and panic of the plain men and women and there has been not only internal
how how really and utterly unfit the plain men and women are to decide on the question of peac
pulled and moved up and down by the plain men and women. Being mere hollow puffed-up puppets,
of the mob, panic of the crowd of the plain men and women communicating itself to and seizing a
and Moses Lump? I answer, because the plain men and women, _even the good honest plain men and
after the Treaty of Portsmouth, the plain men and women everywhere in the country, also raise
se through his Prime Minister to the plain men and women for trying to do his duty as a King t
otect the Emperor of Russia from the plain men and women, from the panic of the crowd of plain
us modern machine in Russia which the plain men and women had helped to create, began to move.
aid, can never be realised. When the plain men and women in the countries of Europe get rid of
Irish question in Great Britain. The plain men and women in Ireland in trying to take into the
l it was this general panic among the plain men and women in Europe seizing and paralysing the
diers and diplomats who have led the plain men and women into this catastrophe, into this hell
mats, who have conducted and led the plain men and women of Europe into this catastrophe, but
Smith, Bobus and Moses Lump tell the plain men and women that they John Smith, Bobus and Moses

Figure 5.13: Concordance of *plain MAN and WOMAN* in CEC

Plain MAN and WOMAN is a way of referring to all ordinary men and women
as one collectivity or group. The meaning of this unique collocation is clear and
acceptable, since it derives from an acceptable usage. It can be seen from this

example that lexical innovations not only occur in individual words but also collocations.

Another unique collocation, *true FACT*, reflects Chinese people's mindset. To an outsider, the use of the word 'true' may seem redundant, yet to a Mainland Chinese user it is an important source of emphasis, that the fact in question is really a fact, as stated in its Chinese equivalent *zhēnxiàng* (真相). This kind of seemingly double meaning is discourse-pragmatically a booster to the Chinese speaker/writer's language, showing sincerity and commitment.

The use of redundant modifiers, where the sense of the modifier is already contained/implied in the word they modify, is very common in Chinese. This is clearly the case of the word *true* in the collocation *true FACT*. Another similar example is *an unfortunate tragedy*, where *unfortunate* is redundant since *tragedy* itself has already included this sense. The exact reason why such a rhetorical strategy is so popular among CE users has yet to be explored, but the tendency of overdoing explicitness or repeating redundant words is believed to be one result of semantic and/or structural influence from Chinese. The concordance of *true FACT* in Figure 5.14 helps illustrate this point (Figure 5.14).



and uncle, which would have been a true fact if I had not been a love child without)
 he advertising language is about a true fact or not, the customer is likely to accep
 his own on Ch'en's behalf. It is a true fact that when a man travels alone he goes a
 essence of a riddle is to express true facts under impossible combinations." I am in
 understanding based on the hidden "true fact" and the logical development of the aut
 here did he come from? Give me the true facts and I will have him brought in. Then w
 t this monstrous distortion of the true facts, but the White Nun ignored her and, tu
 rural areas. l areas . What are the true facts? In 1980 and 1984, the CPC Central Com
 ictims and above all, to teach the true facts of Nanking in Japanese schools. thatâs
 s responsible for ascertaining the true facts of the case, it followed that where th
 friends of Li Chi are seeking the true facts of the crime for which he was wrongful.
 a Ming Collection Ascertaining the true facts of the case, it followed that where th
 de had the effect of bringing the true facts of the case to Hsiang.yuen's notice, an
 As a child, she did not grasp the true facts of the war. As an adult, from her revi
 l being inspected shall report the true facts, provide material relevant to the insp
 the vessel physician must give the true facts. The answers in written form must bear

Figure 5.14: Concordance of *true FACT* in CE

Overdoing explicitness as seen in *true FACT* is a voluntary choice of saying a
 similar meaning twice in order to achieve the effect of emphasis. The reason or
 motivation for semantic redundancy of this kind may be due to the ‘analytical’
 nature of Mandarin whose syllables tend to be morphemic. *Zhēnxiàng* (真相),
 one Chinese translation for *true FACT*, is composed of two Chinese morphemes.
 The first means *true* while the second means *fact*. It may well be the case that



the transference of word-by-word semantic mapping led to a tendency of overdoing explicitness in CE collocations like *true FACT*.

There is an ideological dimension to CE users' collective preference for the collocation *true FACT*, which I think is comparable to the analysis of the ideological significance of the common collocation '*working mother*' (Stubbs 1996, p. 177). Stubbs unpacked the covert, ideologically-loaded meaning of *working mother* in BrE by arguing that as a collocation of *mother*, the word *working* implies that the countless mundane chores that mothers do at home (e.g., cleaning, cooking, etc.) are generally not considered or recognised by society at large as real work, which is why the pre-modifier *working* is needed when reference is made to work that mothers are paid to do. In this sense, the collocation *working mother* seems to reflect a covert value that society places on mothers in general — that what they normally do is not seen as work. Based on this example, Stubbs (1996) drew implications for the general public in UK, and probably elsewhere in the English-speaking world, and pointed out that "if collocations and fixed phrases are repeatedly used as unanalysed units in media discussion and elsewhere, then it is very plausible that people will come to think about things in such terms" (p. 195).



Similarly, in the context of the country with the largest learners and users of English in the world, the collocation *true FACT* may be naturally preferred by CE users for its culturally-conditioned semantic appeal whereby ‘*true fact*’ is differentiated from ‘fact’ per se, reflecting the indigenous Chinese conception that alleged ‘facts’ do not automatically merit or attract the attribute ‘true’. Incidentally, CE users’ collective preference for *true FACT* may serve as an instructive example, which is reminiscent of Stubbs’ analysis of ‘*working mother*’ discussed above, that CE users’ preferred collocational patterns have good potential for revealing deeper ideological roots collectively shared by them, regardless of whether that would amount to deviation – apparent or real – from Standard English norms. Table 5.20 shows further examples of unique innovative collocations in CE:

Table 5.20: More examples of innovative CE collocations

Innovative CE collocations	Tokens in CEC
Peking MAN/Shanghai MAN	136
southern PEOPLE	15
Miao MAN/Han MAN	6
in a timely WAY	5



In sum, apart from overused collocations described in Section 5.2.1.1, innovative/unique collocations were found in 5.2.2.2. These findings, as shown in Tables 5.16 to 5.20, can reasonably expand to a longer list if more similar searches are performed.

However, our point here is to illustrate their existence rather than to come up with an exhaustive list. These findings provide evidence in support of our argument that many CE collocations reflect how Chinese people view the world. Put differently, such culturally conditioned or culture-specific collocations in CE are constitutive of the Chinese way of thinking in regard to their preferred cultural practices, which together make up an integral part of the language identity of Mainland Chinese.

Acceptance towards variant forms of collocaitons of this kind, should therefore, be advocated. CE collocations may thus be crucial in helping people from other cultural background understaning a Chinese way of doing and believing things.



5.2.2 Colligation

In order to identify possible patterns that involve words occurring next to each other we were unaware of before, we need to learn relatively fixed, largely pre-defined sequences of words in more detail. Thus the colligations of these nouns (the target 13 lemmas) are studied next, along with the specific collocational forms of each one.

The significance of this part of the study is to uncover more formulaic patterns that we are unaware of in Section 5.2.1. To conduct our search within workable limits, the modifying patterns of nouns will be chosen as the target, which means the corpus analysis would stick to the L-1 position.

Cross-linguistically, the modifying structures of nouns are very diverse, but the major constituents are nonetheless discernible. According to Leech et al. (2006, p. 71), there are two kinds of modifiers for nouns, namely, premodifiers and postmodifiers. The possible premodifiers and postmodifiers of noun phrases are shown in Tables 5.21 and 5.22:

Table 5.21: Possible premodifiers of noun phrases (Leech et al., 2006, pp. 71-72)

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

Table 5.22: Possible postmodifiers of noun phrases (ibid.)

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

At this stage, words are being analysed in word classes, so their analysis is colligational by nature. The colligation *determiners* + *nouns* is not studied in this thesis because determiners are not comparable in lingua-cultural significance with regard to Chinese.

5.2.2.1 Enumerators as premodifiers in noun phrases

Broadly speaking, an enumerator consists of one or more words denoting a cardinal or ordinal number (e.g., ‘one’, ‘five’, ‘double’, ‘the twelfth’). There is extensive evidence in the CE corpus data showing a clear preference among CE users for using enumerators (Table 5.23).

Table 5.23: Examples of enumerators as premodifiers in CE noun phrases

No.	CE collocations	Chinese	Pinyin
1	One Child Policy	計劃生育政策	jìhuà shēngyù zhèngcè
2	One Country Two System	一國兩制	yīguó liǎngzhì
3	One World One Dream	同一個世界 同一個夢想	tongyīgè shìjiè tongyīgè mèngxiǎng
4	One-China Principle	一個中國原則	yīgè zhōngguó yuánzē
5	Two Guarantees	兩個確保	liǎnggè quèbǎo
6	Double mugging(i.e., snatch and rob, generally referred to in English newspapers in China as ‘Two Robbery’)	雙搶	shuāngqiǎng
7	Two-State Theory	兩個中國	liǎnggè zhōngguó

8	Two-Way Investment	雙向投資	shuāngxiàng tóuzī
9	Three Antis Campaign	三反活動	sānfǎn huódòng
10	Three Direct Links	三通	sāntōng
11	Three Gorges Dam/Hydropower Station	三峽工程/水壩	sānxiá gōngchéng
12	Three Gorges Water Conservation Project	三峽水利工程計劃	sānxiá shuǐlì gōngchéng
13	Three Kingdoms/Period	三國	sānguó
14	Three Public Consumptions	三公	sāngōng
15	Three Represents	三個代表	sāngè dàibiǎo
16	Three Worlds	三界	sānjiè
17	Three-Tier Rural Health Care Service Network	三級保障農村醫療服務系統	sānjí bǎozhàng nóngcūn yīliáo fúwù xìtǒng
18	Four Books	四書	sìshū
19	Four Heavenly Kings	四大天王	sìdà tiānwáng
20	The Four Beauties	四大美人	sìdà měirén
21	The Four Duty Gods	四大金剛	sìdà jīngāng
22	Five Classics	五經	wǔjīng
23	Five Friendlies	福娃	fúwá
24	Five Generations Under One Roof	五代同堂	wǔdài tóngtáng
25	Five Holy/Sacred Mountains	五嶽	wǔyùe
26	The Five Dynasties (Period)	五代	wǔdài
27	The Five Elements Mountain	五行山	wǔxíngshān
28	The Five Principles Of Peaceful Coexistence	和平共處五項原則	héping gòngchú wǔxiàng yuánzē
29	Six Harmonies	六和	liùhé
30	Eight Diagrams/Trigrams	八卦	bāguà
31	Eight Model Plays	八個樣板戲	bāgè yàngbǎnxì
32	Eight Vajrapanis	八金剛	bājīngāng
33	The Eight Allied Forces	八國聯軍	bāguó liánjūn

34	The Eight Arrays	八陣圖	bāzhèntú
35	The Eight Immortals (Crossing The Sea)	八仙(過海)	bāxiān
36	Ten Kingdoms	十國	shíguó
37	The Twelfth Five-Year Plan/Program	第十二個五年計劃	dìshí'èrgè wǔnián jìhuà
38	Eighteen Guardians/Defenders	十八羅漢	shíbā luòhàn

In terms of form, the fact that enumerators function as premodifiers of noun phrases is not surprising. However, their manifestation in CE seems distinct in its own way. In Chinese expressions like those in Table 5.23, the use of numbers symbolises seriousness and carries an unmistakable connotation of authority. Most of the usages are based on the Chinese counterpart and are thus translations in CE. Yet although the Chinese for *One Child Policy* does not include any enumerator, CE speakers still prefer the current way of expressing this idea.

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

5.2.2.2 A preference for noun premodifiers

As we have seen, nouns and adjectives are both common premodifiers of nouns, but their preferred use by CE speakers is another issue to consider. It is found that Chinese speakers tend to use more nouns than adjectives as premodifiers in constructing noun phrases in their social interaction with others in English, and this is supported by corpus investigation in terms of frequency. Collocation results from the KWIC search (span=L-1 left one from the node word) are stored and tagged using the same tagset system as the BNC, namely the CLAWS 5 Tagset (Appendix 1). The tagged words were then sorted in an Excel spreadsheet and counted. After all the nouns and adjectives for the 13 target lemmas are calculated, frequency information per million words is compiled. This standardised frequency information is comparable because they are not influenced by the tokens of each corpus.

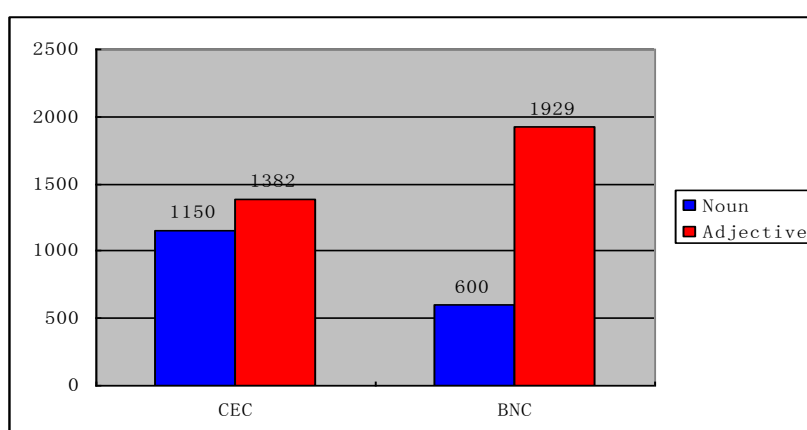


Figure 5.15: Contrasting the proportions of nouns and adjectives as

premodifiers (per million words)

Overuse of the colligation $N + N$, the formation usually referred to as a nominal compound, is an important and useful modification pattern in Mandarin Chinese (and most of the Chinese languages/dialects like Cantonese, Hokkien and Hakka). The most influential description of this structure is provided in Li & Thompson (1981), which lists 21 types of colligation $N + N$ in Chinese (Table 5.24).

Table 5.24: Types of nominal compounds in Mandarin (Li & Thomson, 1981, pp. 49-53)

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



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

*Source: Chinese-English Dictionary (Wu, 2010)

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

We can by no means assume that the 21 types of colligation $N + N$ in Table 5.24 constitutes an exhaustive categorisation of the Chinese $N + N$ pattern. The important thing to note here, however, is that the linguistic preference in Chinese may help us to better understand the more commonly used $N + N$ patterns in CE. CE users link noun and noun together to form a nominal compound, with the effect of designating an object with a name, a productive and creative process which is rooted in their first language, and which by design



expresses their indigenous worldview (e.g. the morpheme-for-morpheme translation of *government organisation* is preferred to the more opaque *bureau*). And, the minor difference in some CE and BrE translations in Table 5.24 is further indicative of differences in CE as opposed to BrE norms (e.g. the high-frequency word *ring*, as in *basketball ring*, is preferred to the more obscure word *hoop*, as in *basketball hoop*).

5.2.2.3 Genitive phrases as premodifiers of nouns

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

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



Two features distinct to CE have emerged, the first of which is the variants and deviations in use of *people's* as a premodifier, as shown in the following examples found in the earlier phase of the investigation (Table 5.25), with three examples of them presented below.

Table 5.25: Examples of using *people's* as a premodifiers only in CEC

people's + N
people's desire
people's thing
people's way
people's cognition
people's awareness

1. Advertisement is a kind of special communicative form and its purpose is to arouse **people's desire** to take action or buy something.
2. As the geographical location and living environment are different, **people's cognition** of the nature is different.
3. Activities of health education in all forms shall be developed to popularise knowledge of old-age health care and enhance the old aged **people's awareness** of self health care.

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

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

One distinctive example is seen in the high-frequency occurrences of the two-word cluster of *people's* + *N* in CEC, with comparison of BNC (Table 5.26).

Table 5.26: Contrasting two-word clusters *people's* + *N* in CEC and BNC

	CEC	BNC
actual tokens	5790	5099
tokens per million	64.3	137.8

We can see briefly in this table that the colligational pattern *people's* + *N* is used more frequent in expressing meaning in CE.

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

The second feature found in this part of the analysis is that CE speakers tend to underuse the colligation *N's + N of*. Examples of this pattern are shown in Figure 5.16 below. One kind of examples is excluded here since they do not fit into the purpose of our investigation; they are institutional names such as *People's Republic of China*, *People's Bank of China*, etc. It is common sense that thousands of these proper names would occur in CE data, yet they do not tell us much about what we hope to prove or disprove.

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

Figure 5.16: Examples of the pattern *people's + N of* in CEC

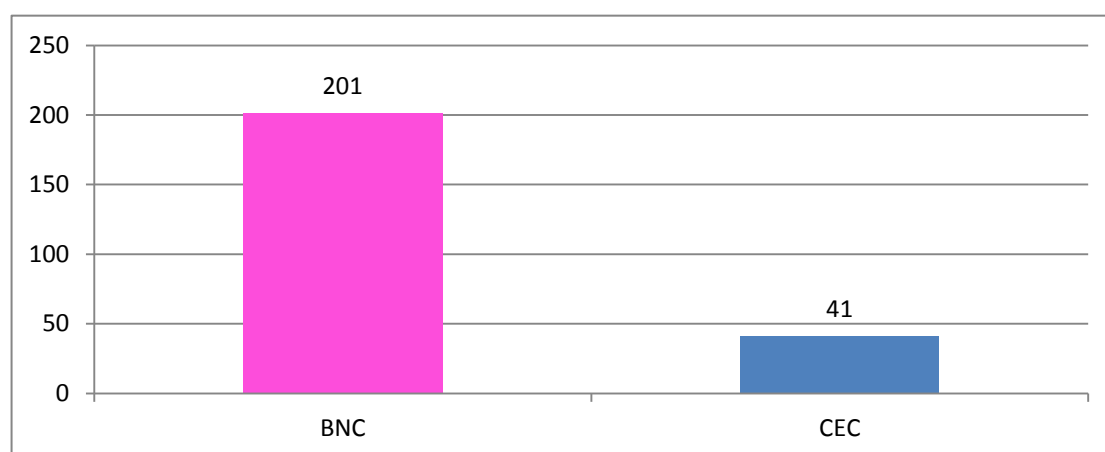


Figure 5.17: Contrasting the proportions of the colligation *N's + N of* in CEC and BNC (frequency per million words)

This colligational structure is constituted by both a premodifier and a postmodifier, and it does not therefore correspond with the Chinese way of expressing ideas. Han Chinese ‘dialects’ or languages are structured according to a fixed word order, namely the modifier-before-head sequence (Tai, 1985; Ho, 1993; Hu, W., 1995).

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

5.2.2.4 Postmodifying prepositional phrases

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



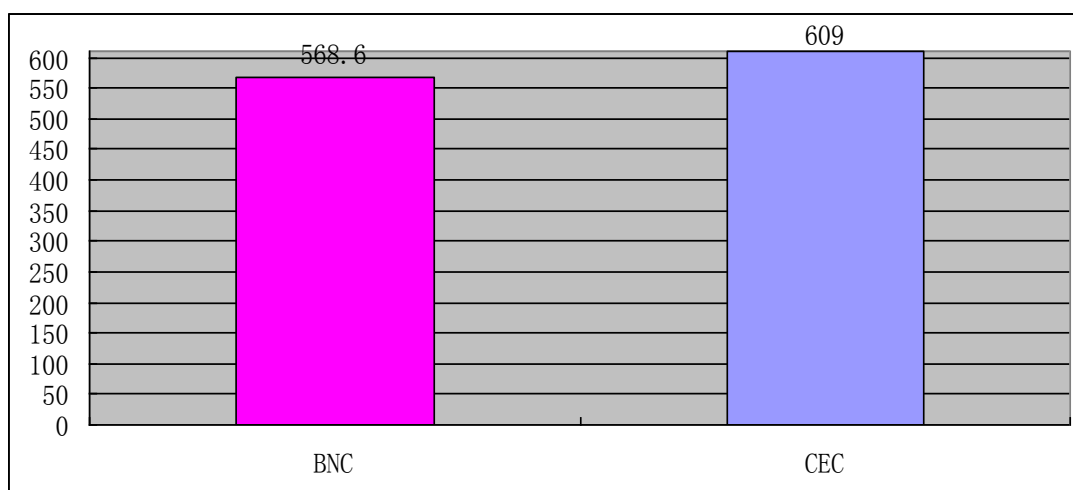


Figure 5.18: Contrasting the colligation *N + of* (frequency per million words)

5.2.2.5 Postmodifying relative clauses

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

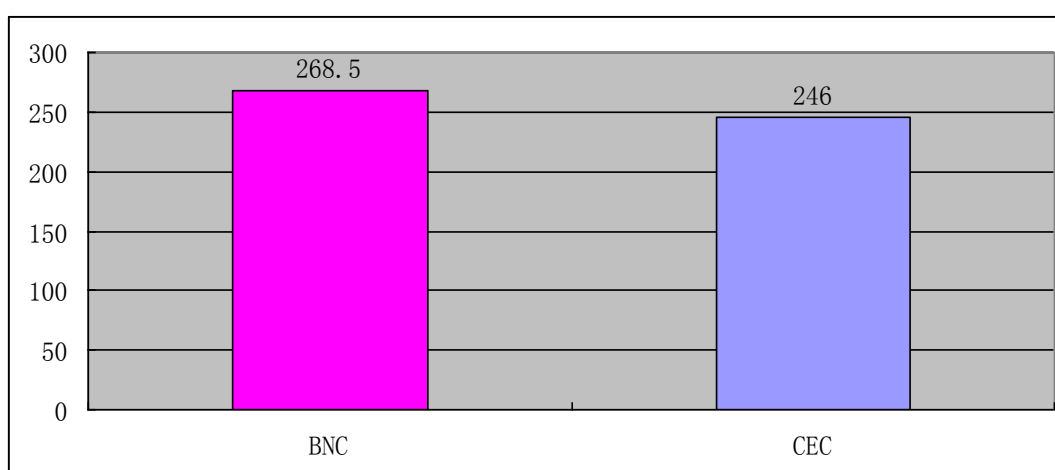


Figure 5.19: Contrasting the colligation *N + that* (frequency per million words)

One possible problem was encountered when studying this pattern. Because in some sentences (e.g., *He would like to give the girl that doll.*), the word ‘that’ functions as a determiner rather than a relative clause marker. Some extra work was done to ensure that Figure 5.19 was not skewed by these sentences. It was found that in the data observed, this problem does not seem to affect the reliability of Figure 5.19. First, the search for this colligation generated a 222,597-token result in BNC and a 60,043-token result in CEC. Then, the actions described in Section 5.1.1 – from initiating to repeating – were taken to observe the first word to the right (R-1) of the pattern *N + that*. Not a single example, as within the limit of manual observation, was found, in the five random pages of KWIC in both corpora.

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



Figures 5.18 and 5.19 show that the two kinds of colligational patterns are used in a similar way in terms of frequency and distribution in CEC and BNC, suggesting that they are largely similar in terms of colligational patterns in the expanding circle variety (CE) and inner circle variety (BrE) being examined.

5.2.2.6 Conclusion of colligation features

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

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



5.2.3 Semantic preference

To shed further light on the distinctive features of CE, it might be interesting at this point to compare the semantic preference of the target lemmas. Given the wide range of word-specific semantic preference, there is a need to narrow down the scope of investigation in order to make the search more manageable.

5.2.3.1 Selection of research focus

In this thesis, the search for semantic preference concerning a particular word is limited to the main verb (Mv., plural: Mv.s hereafter), so the search target would be *Mv.* + *N*. Theoretically speaking, we may find it hard to put all the English verbs into specific groups in an absolute way (e.g., some linguists put *need* under the ‘auxiliary verb’ category while some others classify it as an ‘action verb’).

There are always alternative ways to term them with solid justifications. So we just need to follow one way consistently. In this part of the study, main verbs are distinguished from other auxiliaries as shown in Table 5.27.

Table 5.27: Main verbs in different grammar patterns

Modal	Perfect aspect	Progressive aspect	Passive voice	Main verb ^u
1.				shook ^u
2. might				shake ^u
3.	had			shaken ^u
4.		was		shaking
5.			was	shaken ^u
6. might	have			shaken ^u
7. might		be		shaking ^u

(Leech, et al., 2006, p. 77)

According to this classification, the main verb (Mv.) is obligatory while other auxiliaries are optional elements. Examining the Mv.s related to each target lemma may help us to find out more about patterns revealing CE semantic features. In the examples as shown in Table 5.27, the seven types of verb patterns (e.g., *shook*; *might shake*; *had shaken*, etc.) would all be treated in this part of the study as one Mv. lemma: SHAKE.

The types of Mv. counted in this analysis were retrieved and compared (AntConc setting, span=L-3), and the results indicate that the core Mv.s in both corpora are shared and overlapping.

5.2.3.2 General observation of semantic preference

The Mv. collocates of the target lemmas are shown in Table 5.28.

Table 5.28: Contrasting Mv.s for target lemmas in CEC (alphabetically listed)

Node	Mv.s
AREA	COVER, OCCUPY
CHILD	TEACH, ENCOURAGE, PROTECT, EDUCATE
COUNTRY	LIVE, LEAVE, SPREAD, VISIT
EYE	CLOSE, KEEP, OPEN, CATCH, RAISE
FACT	BE, ACCEPT, BASE, FACE, ILLUSTRATE, ACTIVATE
HOME	GO, RETURN, COME, GET, LEAVE, STAY
MAN	BE, ATTRACT, SAY
PEOPLE	HELP, EMPLOY, ATTRACT, MEET, ENCOURAGE
PROBLEM	HAVE, SOLVE, TACKLE, CAUSE, DEAL, OVERCOME, POSE
QUESTION	ANSWER, ASK, RAISE, POSE, CONSIDER
SYSTEM	USE, INSTALL, DEVELOP, BASE, ADAPT
THING	DO, MAKE, GET, SAY, SEE
WAY	FIND, MAKE, GIVE, GO, COME
DEVELOPMENT	(used as pilot study; see Table 5.29)

5.2.3.3 Semantic preference: PEOPLE, DEVELOPMENT

Though most of the findings suggest that CE is based on BNC norms, some preference for positive Mv.s was also found. By ‘positive’ I mean the word carries a positive connotation. What follows are the statistics for the co-selection of the positive Mv.s –INCREASE, SAVE, IMPROVE–which were found to collocate more with PEOPLE.

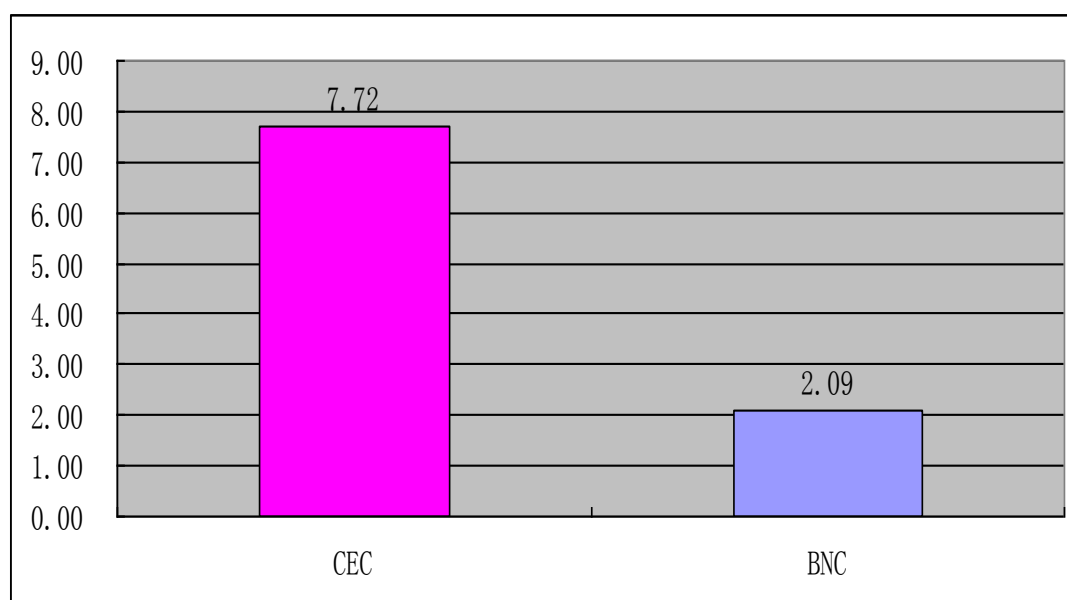


Figure 5.20: INCREASE, SAVE, IMPROVE from L-2 to R-2 of PEOPLE (per million words)

Comparison of other nouns in a similar way may produce more evidence of this preference for positive Mv.s. Some contrastive difference becomes apparent if we examine, for example, the noun DEVELOPMENT (Tables 5.29 and 5.30).

Thus for general/common words (like PEOPLE), CEC and BNC share some similarity rather than difference in the Mv. + N pattern; while for a group of words (like DEVELOPMENT) which tend to be culturally positive to Chinese people, distinct differences in terms of choice of diction and its frequencies are clearly discernible in the Mv. + N pattern (see Tables 5.29 and Table 5.30) (Mv.s with a negator, e.g., *not*, *no*, were not included). For instance, whereas there are 827 tokens of the top-ranked Mv.s of DEVELOPMENT in CEC, there are only 17 in BNC.

Table 5.29: Top 30 Mv.s that collocate with DEVELOPMENT in CEC

(in descending order)

Mv	No. of tokens	Mv	No. of tokens	Mv	No. of tokens
PROMOTE	827	IMPROVE	32	PROCESS	18
SUPPORT	191	EMPHASIZE	27	BOOST	16
ACCELERATE	166	WITNESS	27	FORM	15
ENCOURAGE	67	ACHIEVE	26	INVESTIGATE	9
AFFECT	52	RESTRICT	26	SPUR	9
MAKE	50	ENHANCE	25	REPRESENT	9
SEEK	47	BENEFIT	22	COMBINE	8
INFLUENCE	43	DRIVE	22	DESCRIBE	8
SEE	42	EXPERIENCE	22	EVALUATE	8
STIMULATE	37	STRENGTHEN	21	TRACK	8



Table 5.30: Top 30 Mv.s that collocate with DEVELOPMENT in BNC

(in descending order)

Mv	No. of tokens	Mv	No. of tokens	Mv	No. of tokens
ENCOURAGE	25	<i>INFLUENCE</i>	6	CONSTITUTE	4
SEE	18	<i>REFLECT</i>	6	<u>PREVENT</u>	4
PROMOTE	17	AFFECT	5	<i>SHOW</i>	4
<i>INCLUDE</i>	13	<i>DISCUSS</i>	5	<i>UNDERSTAND</i>	4
<i>INVOLVE</i>	9	<i>EXAMINE</i>	5	<i>COMPLETE</i>	3
<i>ALLOW</i>	7	FOSTER	5	<i>CONSIDER</i>	3
<i>MONITOR</i>	7	REQUIRE	5	<i>COVER</i>	3
<i>TRACE</i>	7	SUPPORT	5	<i>EXPLAIN</i>	3
<i>DESCRIBE</i>	6	<i>WATCH</i>	5	FACILITATE	3
<i>FOLLOW</i>	6	ASSIST	4	UNDERGO	3

Comparing and contrasting Tables 5.29 and 5.30, we may find that there are

- a) Shared items which are used in both varieties;
- b) Distinct verbs used by BrE (italicised); and
- c) Negative verb PREVENT (underlined).

Our purpose is to demonstrate that for common and general nouns like the 13 target lemmas, patterns in terms of Mv. + noun are shared between the two corpora, and that CE usage does not take on specific colours of the Chinese

language. As collocates, ‘PROMOTE + DEVELOPMENT’, ‘SUPPORT + DEVELOPMENT’, and ‘ACCELERATE + DEVELOPMENT’ occur 1,184 times (32 times per million words) totally in CEC, reflecting a relatively positive attitude towards DEVELOPMENT in a sense, where as the top-ranked Mv. in BNC (ENCOURAGE) only occurs 25 times (0.27 times per million words). In other words, a heavier load is assigned by CE users to the Mv.s with positive connotations to collocate with DEVELOPMENT, resulting in a stronger sense of preference for its use.

5.2.3.4 Conclusion of semantic preference

For lemmas which do have special connotations in Chinese culture, features of semantic preference may be found through other means. Another way of viewing this phenomenon is perhaps to see it as a colour continuum, where more or less distinct colours flow into each other. The preferences suggested by these Mv. collocations could be visualised in terms of a ‘positive-neutral-negative’ cline, as shown in the following diagram:

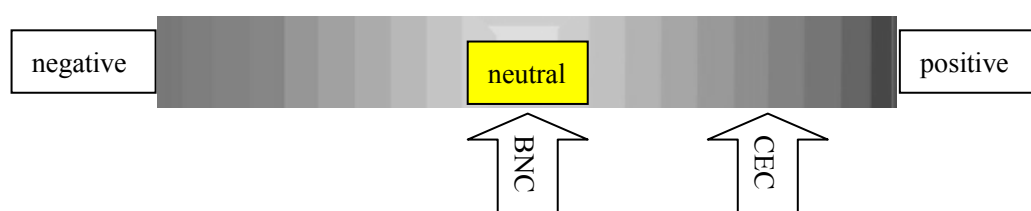


Figure 5.21: Contrasting semantic preferences along a colour chart

These results suggest that subsequent research may consider investigating lexical items that are overtly Chinese (e.g., *hutong*, *guanxi*), and should shift from the surface to the deeper layers of meaning which are valued or of particular interest in Chinese culture.

To illustrate this point, one more example is relevant here. For Hakka people (an ethnolinguistic group that numbers about 30 million around the world, www.ethnologue.com/language/hak/view/***EDITION***), one common way of making a compliment concerning someone's body shape/fitness is to say that 'you are fatter than before'. In Hakka culture, the word 'fat' bears special meaning which is comparable to 'fit, strength, and a prosperous living condition'. So it is presented with positive connotations. I feel extremely shocked at the frequent encounter with people (though communicating in Cantonese or Mandarin) from Hakka-speaking regions saying things like "you are fatter than before!"

Similar cases are reported by Günthner (2002) in her contrastive study of the strategies of behavior reflected in German-Chinese verbal communication. People from different cultures would draw their own inferences about 'right' and 'wrong' behaviour, which are characteristic of their own cultural norm. For example, Günthner illustrates a great difference between western culture and



Chinese culture in regard to the issue of ‘being active or not’. Western culture tends to favour ‘active’ and shows a tendency to believe that “one should always be active”, while Chinese people tend to adhere to a deeply-rooted norm whereby “one has to remain passive and quiet” (Günthner, 2002, p. 367).

The culture-specific social values about ‘fat’ and ‘quiet’ are examples of a “discourse system” (Scollon & Scollon, 1995) which could be seen as the root cause of why people tend to feel socially obliged to produce such speech acts, even though in theory many others could be considered in their social interaction with others. CE features such as these, in which linguistic choices take on cultural colours, manifesting in whole utterances rather than single words, collocations or grammatical categories (i.e., colligation), are currently not received enough attention in research, and deserve more careful attention.

5.2.4 Semantic prosody

Semantic prosody (Louw, 1993) refers to the idea that words tend to collocate with specific semantic groups as well as individual words. For example, the word *hair* may collocate with words denoting length (e.g., *long*, *short*) or colour (e.g., *red*, *blonde*, *black*). So far, a clear conceptual distinction between semantic preference and prosody is not yet fully and satisfactorily resolved



since such things as “aura of meaning” (Louw, 1993, p. 157) or “spreading of connotational colouring” (Partington, 1998, p. 68) are difficult to identify. The current situation is that disagreement still “cluttered the debate since the phenomenon [semantic prosody] was first brought to light” (Louw & Chateau, 2010, p. 756). However, “the importance of authentic examples” (ibid.) is the key to unveil semantic prosody, so examination of concordances generally helps to reveal this kind of extended meaning of a word by analysing the semantic properties of its collocates.

In Section 3.2, the notion of semantic prosody as used in this study has been defined. It is the driving force behind the selection of the core and the other co-selections comprising the lexical item which is established according to their attitudinal attributes, that is, on the basis of whether they have a positive, negative or neutral association.

5.2.4.1 Selection of research focus

As pointed out by Sinclair, “semantic prosody is attitudinal” (2004, p. 34), so our target is restricted to adjectives (hereafter Adj.s) that modify the node lemmas under investigation. The concordances suggest that there is greater consistency of patterning to the left of the collocation than to the right, so we



move our focus to the left 1 (L-1) position. The findings related to semantic prosody in CE will be discussed in three parts:

- a) General observations;
- b) Deviations in semantic prosody; and
- c) Semantic prosody of the target lemma SPIRIT.

5.2.4.2 General observations of semantic prosody

Appendix 2 shows the 50 most frequent adjectives in the L-1 position associated with each of the 13 target lemmas in this study. A shorter list presenting the shared adjectives in both CEC and BNC is shown in Table 5.31. Listed in Table 5.31 are the shared adjectives among the top ten in both corpora.

Based on careful examination of this table and appendix 2, particularly by comparing the adjectives in CEC list with the same lemmas in BNC, it can be readily concluded that the semantic prosody for those general nouns is, on the whole, similar.



Table 5.31: Contrasting Adj.s for target lemmas in CEC and BNC

Node	Adj
AREA	whole, western, other, foreign, European, different, developed
CHILD	young, other, little, small
COUNTRY	other, developed, foreign, different, western, whole, European
EYE	blue, black, green, brown, small, big
FACT	important, historical, social
HOME	ideal, new, old, spiritual
MAN	old, young, other, good, dead
PEOPLE	young, old, other, disabled
PROBLEM	serious, environmental, social, main
QUESTION	following, simple, open
SYSTEM	legal, economic, social, political, new
THING	other, only, important, good
WAY	different, long, effective, other, only, new

Above are the shared adjectives in both corpora within top-ten ranking. General adjectives are used. As for the longer list in appendix 2, this impression is confirmed. Take PEOPLE as an example, the top three adjectives in the CEC are *young*, *other* and *Chinese*, while in the BNC they are *young*, *other* and *old*.

The top three adjectives modifying EYE are *own*, *blue* and *black* in the CEC,

and *blue*, *brown* and *grey* in the BNC, all of which are associated with the colours of human eyes. Thus the view that high-frequency nouns tend to share similar semantic prosody seems to be supported by the results here, which perhaps makes sense, since in communication with people from other cultures one tends to use culturally neutral terms, probably to avoid confusion. In other words, the core norms are shared.

5.2.4.3 Semantic prosody: PEOPLE, CHILD, SPIRIT

Besides the core features, there are also some deviations, reflecting the attitudes of users towards the node lemma words. The findings for PEOPLE and CHILD are interesting in this respect. The adjective *Chinese* (as in the collocation *Chinese PEOPLE*) ranks number one in terms of frequency as an L-1 collocates, which is not surprising.

However, two other adjectives, *common* and *ordinary* are also useful indicators of the emphasis in CE on a ‘grass roots’ sense or quality in these nouns, which has been discussed in an earlier part of the thesis (see Figure 5.5; cf. Section 5.2.1.1):

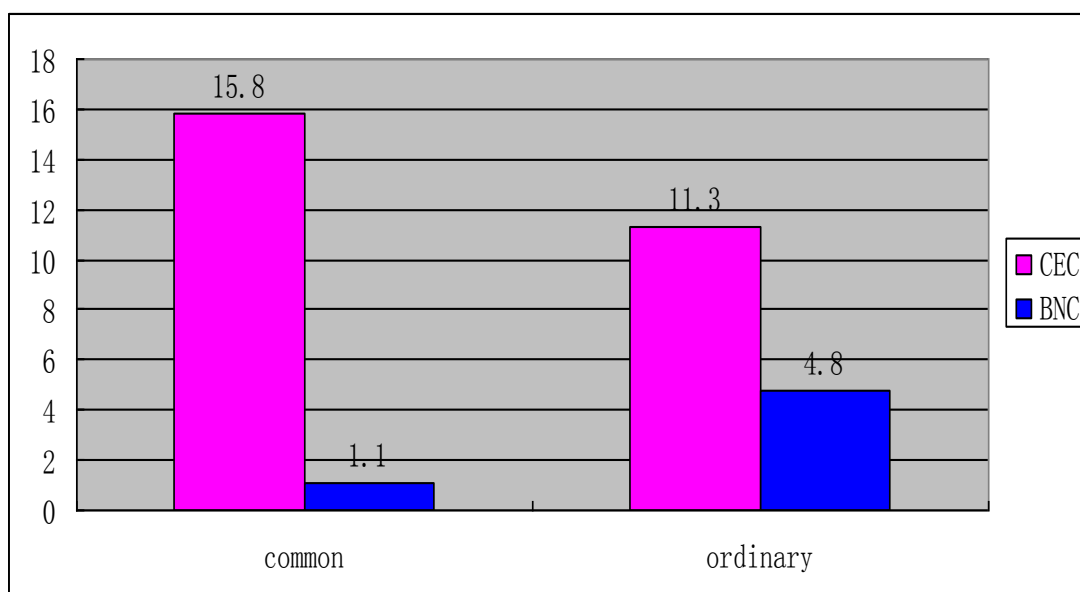


Figure 5.22: Contrasting *common PEOPLE* and *ordinary PEOPLE* (per million words)

It is important to understand that the meaning of *common people* has positive connotations in Chinese culture, with a feeling of closeness or established intimacy that permeates the use of this word in CE. By using *common people*, CE users are referring to 老百姓 (lǎobǎixìng) and 人民 (rénmín), which all have “good prosody” (Louw, 1993, p. 171). Dictionary entry (Chinese-English Dictionary) provides English for 老百姓 as ‘*common people*’ (Wei, 2007, p. 721). Thus the collocation *common PEOPLE* in CE takes on positive colouring mostly due to the fact that the Chinese for it is used in a favourable way. The following examples are taken from CEC data to show how this term is used in a variety of contexts.

1. Its forging technology was kept a secret and the burners had never even been seen by **common people**.
2. However it started, the **common people** were soon following the example of their rulers and the observance of the Lantern Festival spread all over the country. Eventually it came to mark the end of the New Year season.
3. **Common people** write their blogs to share their thoughts or experience with friends or even more blog-visitors. Some have become famous overnight.
4. The aim was to carry forward revolution smoothly, which was in the interests of the **common people**.
5. In such a society, only a few political figures have freedom and democracy, while the **common people** lack even adequate food and clothing and barely survive, let alone pursue freedom and democratic rights.

The collocation *common PEOPLE*, as presented above, is believed to have a special aura distinctive to CE.

Interesting semantic choices are also found in the collocates of the word CHILD. It was found that the adjectives modifying CHILD are mostly shared by both the CEC and the BNC, as summarised in Table 5.32.



Table 5.32: Contrasting the adjectives pre-modifying CHILD in CEC
(alphabetically listed)

	Semantic prosody	Adjective
group 1	negative quality	bad, noisy, spoilt
group 2	pitiful state	autistic, blind, dead, disabled, handicapped homeless, illegitimate, innocent, poor, retarded, sick
group 3	position in the family	first, fourth, only, own, second, third
group 4	descriptive	foreign, future, little, local, newborn, old, rural school-age, small, traditional, unborn, young
group 5	positive quality	good, happy, lovable, lovely

The main attitudinal difference is found in the category of relative independence, as shown in Table 5.33 below:

Table 5.33: Contrasting the frequencies of occurrence of *individual CHILD*

	Tokens in BNC	Token in CEC
individual CHILD	116	1

Attitudes toward children may be reflected in the use of this collocation since the semantic prosody of CHILD is embedded in culture-specific assumptions.

Chinese culture holds a different view toward children when compared to the

western cultures, one that goes contrary to the view of children as individuals. In China, it is considered a must for children to be affiliated with their adult dependents, typically their parents or next of kin (sometimes for life, and regardless of age). So CE reveals a set attitudinal perspective towards CHILD which is far removed from *individual* (Table 5.33).

Regarding the traditional Chinese view toward children, one instructive example is the traditional classic text used for teaching children: *Rules and Disciplines for Children* (Dìzǐguī《弟子規》, Y. X. Li, ca. 1661-1722), where the normative standards for being a good youngster are spelled out. The title of the book means: *Standards for being a Good Child*. It was believed to be written in the Qing Dynasty during the reign of the *Kangxi Emperor* (1662-1722 AD).

Written in three-character verses, it is a primer highlighting the ancient teaching of Confucius that emphasises the basic requisites for being a good pupil/child and, in a broader sense, serves as guidelines for good behaviour which is crucial for living in harmony with others.

The book is composed of seven chapters with each chapter listing one duty that a good child should follow in life:



Table 5.34: Contents of *Rules and Disciplines for Children*

Chapter	Chinese	Gloss in English
1	孝	On Being Filial
2	悌	On Practicing True Brotherhood
3	謹	On Being Cautious
4	信	On Being Honest
5	泛愛眾	On Loving All Living Beings Equally
6	親仁	On Being Close to and Learning from People of Virtue and Compassion
7	學文	On Studying Whenever You Can

Listed below are 10 normative, good behaviours expected of children out of the 360 sentences in this book:

- a) 父母教須敬聽 ‘When parents teach, one must listen respectfully’;
- b) 父母責須順承 ‘When parents reprimand, one must accept’;
- c) 親所好力為具 ‘What parents like, make an effort to possess’;
- d) 親所惡謹為去 ‘What parents despise, carefully get rid of’;
- e) 親愛我孝何難 ‘When parents love me, it is easy to be good to parents’;
- f) 親憎我孝方賢 ‘When parents detest me, then it’s truly virtuous to be good to parents’;

- g) 親有過諫使更 ‘When parents are doing something that’s wrong, counsel them to make them change’;
- h) 怡吾色柔吾聲 ‘Make my expression pleasant, and soften my voice’;
- i) 諫不入悅復諫 ‘If the counsel is not accepted, wait until parents are happy then counsel them again’;
- j) 號泣隨撻無怨 ‘Weeping and wailing should follow, and beatings are borne without complaint’.

(based on Taylor & Legge (Translator), 2011, author’s selection)

Key words for a good child include ‘dutiful’, ‘obedient’, ‘disciplined’, ‘tolerant’, which represent a typical Confucian tradition, one that different from the Western view regarding what a child should do. One implication for CE is that such cultural values are naturally reflected in its use, as shown in the collocates of PEOPLE and CHILD, among others. To put it differently, we may say that these differences should not be seen as improper usage. On the contrary, only by accepting the semantic preference of CE can we begin to recognise and respect Chinese culture.

One current example that may show *individual* and *child* are not connected is a report by Liao (2013). She interviewed 130 freshmen in universities in Chongqing City, Mainland China, on the first day of school and found that 90

percent of them were ‘accompanied/escorted’ by parents (even grandparents in some cases). Most parents took a plane or train all the way to the university together with their children, just for the sake of helping them with chores like carrying luggage and registration. After these, they stayed at or around the universities to accompany their children for days. All the hotels were fully booked. As this happened every year, Chongqing University had prepared 200 tents in the gymnasium for the parents to pass the nights (Liao, *Chongqing Economic Times*, 17th September, 2013). This may seem surprising in other cultures, if a person aged 20 (around this age when they enter university) would still be taken care of or even nursed like a child, but it appears to be quite normal in China.

At the pilot phase of this study, the word SPIRIT (meaning ‘alcoholic’ excluded) also came under focus. This word lemma is not in the high-frequency list of words in general English. Rather, it is an author-nominated item for scrutiny. This is motivated by my personal intercultural communication experience, where tremendous differences between East and West regarding how we look at this word were detected.

The attitudinal difference shown in usages of this word may deviate from the way people understand it. SPIRIT (jīngshén 精神) is one high-frequency word



appearing in various Chinese slogans (e.g., 雷鋒精神, the spirit of Leifeng; see Table 5.35). It is thus interesting to compare the findings for the lemma SPIRIT with those so far established.

Table 5.35: Common Chinese collocations of SPIRIT and 精神 (jīngshén)

CE	Chinese	Pinyin
the spirit of reform	改革精神	<i>gǎigé jīngshén</i>
the scientific spirit	科學精神	<i>kēxué jīngshén</i>
the spirit of the Fifth Plenary Session of the Party's Fifteenth Central Committee	十五屆五中 全會精神	<i>shíwujìè wúzhōng quánhuì jīngshén</i>
national spirit	民族精神	<i>mínzú jīngshén</i>
innovative spirit	創新精神	<i>chuàngxīn jīngshén</i>
spirit of the talk	講話精神	<i>jiǎnghuà jīngshén</i>

These examples of collocations in Chinese show that unusual subjects (e.g., *reform*, *Committee*, *talk*) could collocate with SPIRIT. However, in search of the aura/ semantic prosody of CE usages of this word, we need to further examine the adjectives modifying this lemma in authentic examples, as shown below (Table 5.36).

Table 5.36: Contrasting Adj.s pre-modifying SPIRIT (alphabetical order)

	CEC	BNC
semantic prosody	showing appraisal, admiration	describing facts and conditions
Adj.	adventurous angelic artistic collective Confucian creative enterprising entrepreneurial fragrant free generous glorious good guardian heroic high honored humanistic independent indomitable inner innovative military national	adventurous bad certain competitive corporate critical democratic different divine entrepreneurial essential evil festive free generous genuine good great guiding holy independent indomitable inner kindred

	new	liberal
	noble	moving
	Olympic	national
	optimistic	new
	patriotic	old
	pioneering	pioneering
	proud	public
	public	pure
	reasonable	real
	rebellious	restless
	revolutionary	revolutionary
	scientific	scientific
	selfless	similar
	sporting	surgical
	strong	troubled
	team	true
	true	universal
	youthful	white

From Table 5.36, it can be seen that SPIRIT is a representative from a group of words that have a distinctive semantic prosody in CE, where a CE speaker might expect to be differ noticeably from BrE usage. Other similar examples may be found and added to the present list. The examples below, found in *Report on the Work of the Government (2000-2013)*, could shed more light on the ‘colour’ of the word– ‘spirit of reform’, ‘unyielding spirit’, and ‘pioneering spirit’:

1. All departments under the State Council and local governments at all levels should make it a priority this year to improve their administration in the spirit of reform.

Chinese version: 今年，国务院各部门和地方各级政府都要以改革的精神，突出抓好管理。

2. This hard struggle against the earthquake produced uncountable and unyielding spirit of the Chinese people, thus writing a heroic chapter in the history of the nation.

Chinese version: 这场艰苦卓绝的抗震救灾斗争，涌现出无数感天动地、可歌可泣的英雄事迹，充分展现了中国人民不屈不挠、自强不息的伟大民族精神，谱写了气壮山河的壮丽篇章。

3. Over the past five years, the people of all ethnic groups worked hard together, blazed new trails with a pioneering spirit, and obtained great achievements in all fields of endeavor under the leadership of the Communist Party of China.

Chinese version: 过去五年，全国各族人民在中国共产党领导下，以团结奋斗，开拓创新精神为支持，在各个方面取得了重大成就。

Summarising this section, it has been found that in many cases, the semantic prosodies of many words in inner circle and expanding circle English varieties are similar. These phenomena reflect the fact that high-frequency words play a



common role in the English language as the basis for mutual understanding. Yet we also see that some prosodically loaded forms in CE tend to crowd around words such as PEOPLE and CHILD, revealing that there are well-established, culturally conditioned and internalized attitudes among Chinese speakers towards certain aspects of human life and society. These semantic prosodies do not, however, interfere with the core norms of communication, and thus should be considered acceptable.

5.2.4.4 Conclusion of semantic prosody

The results reported in this chapter capture some aspects of the collocational features of CE, and can make a potentially significant contribution to the study of World Englishes, where the identification of patterns of word usage in CE – and possibly other new varieties of English – is an important issue. It is hoped that the analytical procedures outlined in this chapter could help refine the description of CE as a developing variety within the WE family. Alternatively, one may as well see the findings as a set of working hypotheses rather than established CE features, which may serve as evidence shedding light on the direction of further research in the field. Discussion of the research questions and hypotheses will proceed in the next chapter.



5.3 Summary of chapter

This part of the thesis illustrates and presents how CE deviates from BrE in terms of collocational features. The step-by-step procedure was narrowed down with justification. The focus of study was placed on high-frequency nouns as these lemmas constitute more important weight in daily communication activities in terms of representativeness. After comparison and contrast, we have targeted 13 high-frequency nouns as the node words for investigation. Then, the four parameters of meanings (described in Chapter 3) for these nodes were studied. Findings of each type of collocational parameter were illustrated by comparison and contrast following the rationale described in Chapter 4. We have illustrated these findings with examples and KWIC to show the different grammatical constructions in which the node words can occur. Conclusions and statements addressing the research questions, as raised in Chapter 2, will be presented in Chapter 6.



CHAPTER 6

DISCUSSION

A systematic investigation of collocational features of CE has been carried out, and the results, obtained through the quantitative analysis of CE features supported by some illustrations from the qualitative data, were presented in Chapter 5. This chapter discusses the findings with the aim of explaining the reasons behind both the common collocational properties of CE, and the collocational and structural variations, while at the same time hoping to provide informed answers to the research questions set in Chapter 2. A number of writers, e.g., Xu (2010), C. Gao (2009), and Yu (2009), have shown that Mainland Chinese have a particular way of expressing ideas, and it is likely that further evidence will accrue in support of these observations as time progresses. Furthermore, as we saw in Chapter 2, since the 1980s, there has been a heated debate among Mainland Chinese scholars about whether CE exists as a new variety or not, with each side presenting weighty arguments, albeit with little empirical evidence, quantitative or qualitative. As mentioned in Chapter 5, CE collocational patterns may be observed in texts authored by Mainland Chinese academics across a wide range of disciplines. The results as shown in Section 5.2 provide us with fairly good examples of CE collocations (e.g., the



high-frequency use of *all-round WAY*, the unique use of *a scientific outlook on development*), the components of which are frequently used, but put together, the CE collocations are non-existent in BrE. Based on the analysis supported by illustrations in Sections 5.2.1 to 5.2.4, several conclusions may be drawn from these quantitative results where CE collocational patterns have been identified. In the following sections, some of these patterns will be discussed with more contextualized examples for illustration.

6.1 Existence of collocational features of CE

As stated in Chapter 1, the first research question is: “Are there distinctive collocational features in the English used by Chinese?”, which is about the existence of collocational features of CE. In Chapter 5, we have seen deviations from BrE that emerge in the form of co-occurrences of words (e.g., *in an all-round way*), or word categories (e.g., preference in using *N + N*), which are independent of errors and culture-specific lexis (e.g., *hutong*, *hukou*) (cf. Tables 5.6-5.7, Sections 5.2.1.1-5.2.1.2). Now we will first discuss the existence of collocational features of CE, by using KWIC to identify three examples of innovative collocations (cf. Tables 5.17 and 5.20): ‘(in an) all-round WAY’ (Figure 6.1), ‘eco-economic SYSTEM’ (Figure 6.2), and ‘Shanghai MAN’ (Figure 6.3).



KWIC	
nt of the armed forces, the PLA has adopted in an all-round way a mode of procurement of military materials	
isks and measures for our country to advance in an all-round way the administration by law in the next ten y	
nd its ecological restoration work. Advance in an all-round way the construction of comprehensive recycling	
In 1982, the Constitution was amended again in an all-round way, and the amendments were adopted at the Fif	
thletic ability of the coaches and athletes in an all-round way. Strengthen the organization and management.	
ent of exercise of regional ethnic autonomy in an all-round way. From then on, the Tibetan people were enti	
A: Achieving moderate prosperity in China in an all-round way is a goal set by the Party's Sixteenth Nation	
on for building a well-off society in China in an all-round way. Our great motherland already stands at a h	
ity and social progress in the capital city in an all-round way. (Two) Strategic Goal Provide all-round top	
expedite the country's social construction in an all-round way in order to clear the path for sustained and	
aidance of "Students are expected to develop in an all-round way - morally, intellectually and physically,"	
d education to ensure that students develop in an all-round way. We will accelerate reform and innovation in	
hasis to ensure that students would develop in an all-round way, morally, intellectually, physically and ae	
ion and other social undertakings developed in an all-round way. Technological innovation and translation o	
e and Art Are Being Inherited and Developed in an All-Round Way The Central People's Government and the Peop	
th, sports and other undertakings developed in an all-round way. The success of the Shenzhou VI manned spac	
g, and radio, film and television developed in an all-round way and continued to bring out fine works in la	
ion and other social undertakings developed in an all-round way. Technological innovation and translation o	
ip with South Asian countries is developing in an all-round way. President Hu's visits to India and Pakista	
out the scientific viewpoint of development in an all-round way, establishing a harmonious socialist societ	
inuously pushes forward women's development in an all-round way. The Beijing Declaration and the Platform f	
a closed society to high-speed development in an all-round way in opening-up, cooperation and competition.	
inuously pushes forward women's development in an all-round way. The Beijing Declaration and the Platform f	
introduced this year to develop the economy in an all-round way. A large number of the current NPC deputies	
lopment; to promote quality-based education in an all-round way; to coordinate educational resources; to de	

Figure 6.1: Sample KWIC of *(in an) all-round WAY* in CEC

KWIC

activity of developing economy just goes along in eco-economic system, which must affect the running of natural ecological supply and economic demand balance in eco-economic system and the one of supply and demand balance in China makes us understand human domination in eco-economic system is not always proper. Concretely, when it leads to wrong exertion of human domination in eco-economic system. Therefore, more and more serious ecological problems. For example, due to the feature of energy flow in eco-economic system, energy not used emits into air in the form of waste heat between material circulation and energy flow in eco-economic system. Therefore, according to the principles of the features of energy flow and information flow in eco-economic system. For example, due to the feature of energy flow, while, when knowing human position and function in eco-economic system, man should understand his bidirectional relationship with nature, to fully exert human dominative function in eco-economic system is very important. However, the long-term ecological balance. The feature of dualism of human position in eco-economic system points out that when using the natural resources, substance and energy transfer and transformation in eco-economic system shows as substance and energy flow in eco-industrial system. It refers to an industrial eco-economic system, in which resources and energies have been used for forming harmonious and developed industrial eco-economic system, establishing perfect regional, urban and rural substantial conversion and energy flow in industrial eco-economic system. That is to say, in the process of substantial conversion and energy flow in industrial eco-economic system, partial substance and energy have been converted into industrial eco-economics; Eco-industry; Industrial eco-economic system; New type of industrialization road 1. The structure, function and running law of industrial eco-economic system that is formed with industrial economic system and highly efficient running of industrial eco-economic system, which is combined with industrial economic system is in favor of improving structure of industrial eco-economic system and the structure of eco-industry, etc. The resource utilization efficiency of the integrated eco-economic system by relating the area of footprint to the area of the system. At this time, man is not only one of the elements of eco-economic system but also a dominator of it. The reason is that the dominator of it. The reason is that the essential of eco-economic system is human eco-system rather than purely natural system is an accumulation of small and big regional eco-economic system as well. In the ecological economic system, the level of consumers into one via food chain. In urban eco-economic system, various eco-economic elements are combined into chains - investment and production chains in urban eco-economic system respectively. The thinking way and method

Figure 6.2: Sample KWIC of *eco-economic SYSTEM* in CEC



ng outrageous were to occur affecting his life, a Shanghai man, initially, is not likely to fight against i
 neral lack of "pioneering spirit" was noted among Shanghai men, whose average income, in their late 20s, is
 alculate about only several yuan in small bills." Shanghai men are also accused of lacking courage to stand
 g or buried alive in Shanghai can be the blessed "Shanghai Man." We poor journalists, for example, have cer
 oison. To some people, the advan-tages enjoyed by Shanghai men are what they are notorious for. To be neat
 s some negative overtones for most men in China. "Shanghai man tends to be very self-protective. They are u
 ly, is not likely to fight against it. This gives Shanghai man another bad name as easily-intimidated coward
 about it. Yes, there are obvious shortcomings in Shanghai men, but they have arisen with the commercialisa
 washing and ironing their only shirt every night. Shanghai men also use perfume, and they use it with great
 from his career. Women enjoy the attentiveness of Shanghai men but can't stand their being fussy and "picky
 one's looks is too much. The typical neatness of Shanghai men has gained them a reputation for lacking mas
 shoes and his hair slicked down with lots of oil. Shanghai men are smaller in average size than men from th
 s in personal affairs as well as in business. One Shanghai man has written to a local newspaper, the "White
 es, which is comparatively boring and repetitive. Shanghai men are known throughout China as model husbands
 d with the establishment of the city of Shanghai. Shanghai men accepted the fact from the very beginning. I
 erparts wipe off sweat with their loose T-shirts, Shanghai men are carefully washing off any trace of sweat
 nd for affectation. A caricature of stereotypical Shanghai man would show him in shining shoes and his hair
 ure too. In current Chinese literature, the term "Shanghai Man" has long been used as the synonym for a Bab
 r weird to see men busy cooking and washing. That Shanghai man rushes home as soon as work is finished arou
 n though they have little body odour to hide. The Shanghai man is known to be good at housework. This is an
 mple life and had a lower standard of living, The Shanghai man would be up washing and ironing their only s
 able and clothed in lurid beauty. Apropos of the "Shanghai Man' Walking along Nanking Road in a sunless Sun
 d all's right with the world! Like the poet, the "Shanghai Man" is born, not made. Not everybody living or
 said a local woman. Call it shrewdness or thrift, Shanghai men's habit of haggling over every little detail
 es and to small details but lack a "macro view". "Shanghai man tends to pick on others but pay little atten

Figure 6.3: Sample KWIC of *Shanghai MAN* in CEC.

As can be seen in the first KWIC concordance (Figure 6.1), the collocational pattern ‘*in an all-round WAY*’ is used consistently by a large number of CE authors belonging to and working in different academic disciplines (e.g., social science, education, world affairs). One important point that should be reminded of is that the collocation *all-round WAY* is not found in BNC, although separately, these two words are used in other patterns, for example, *all-round player*, *best way*, but not in the pattern *all-round WAY*. As shown in the KWIC concordance lines in Figure 6.1, this collocation is embedded in longer patterns such as ‘*an all-round WAY, in an all-round WAY, DEVELOP in an all-round WAY*’, etc. These longer patterns provide strong evidence of the existence of the CE-specific collocation *all-round WAY*, thereby giving an affirmative answer to research question one.

A similar finding was obtained in the second KWIC concordance lines ‘*eco-economic SYSTEM*’ (Figure 6.2), which presents a collocational pattern that is not used in BNC, but systematically used by CE users in their writing on various topics. However, there is one more important implication in this example: we can find that not only the ways words combine with each other are different from BrE, but the grammatical structure of this collocational pattern also deviates from the usual pattern. For instance, most tokens of ‘*eco-economic SYSTEM*’ are not preceded by a determiner, and yet meaning is



not adversely affected in any way. By contrasting corpus search results in BNC, we found that a determiner (e.g., *a, the, this*) or a pronoun (e.g., *his, her, our*) are obligatory for the structure ‘*in + ? + SYSTEM*’. But this grammatical rule is not adhered to in this CE collocation pattern. There is thus room for arguing that CE has evolved its own norms in regard to the use of the collocational pattern ‘*in eco-economic SYSTEM*’, in that a preceding determiner is dispreferred by CE users.

The third KWIC shows that ‘*Shanghai MAN*’ is a concept coined to refer to male Shanghainese with specific characteristics. The interesting point lies in the ‘aura’ of this coinage, that is, the embedded semantic prosody or connotation. For instance, ‘*not likely to fight against something*’ in line one, ‘*lack of pioneering spirit*’ in the second line, ‘*lacking courage*’ in the third line, are among many attributes of the ‘spineless’ character of Shanghai men, which is a seemingly well-known trait in the study of Chinese culture. As we read through the concordance lines, we may find other descriptive key words attributed to the notion of *Shanghai MAN* (e.g., *perfume, cooking, haggling over every little detail*, etc.). This collocation is thus conceptually more complex than the above two KWICs. To understand it more fully, one needs to know the full range of prosodic meanings embedded in Chinese culture’s ‘routine’ understanding of *Shanghai MAN* (in contrast to the ‘masculinity’ for men in Peking, i.e. *yémēn* 爺).



們, a usual descriptive word for men from Shanghai region is *xiǎonánrén* 小男人, literally ‘little male person’). Below is an instructive contextualised example.

Many years ago, some readers challenged Yu Qiuyu, a famous writer in Mainland China, of not having a balanced identity between being a scholar and a writer (in a journal *Free Forum of Literature*; Retrieved from <http://www.tjswl.cn/2009/4-28/16142239714.html>). In response to this challenge, Yu reacted strongly by composing a preface of 25 thousand words in his book 《山居筆記》 (*Shānjūbǐjì*, ‘Notes from my hillside cottage’). People then attributed this to Yu being a ‘Shanghai man’, which includes such attributes as being too stingy, narrow-minded, trivial and over-sensitive (www.news.sina.com.cn/living/9905/051101.html). To Mainland Chinese people, ‘*Shanghai man*’ tends to be associated with being trivial in household chores (e.g., knitting, cooking), and shrewd in the way they spend money.

The content analysis of the expanded collocational patterns and semantic prosody of the three CE-specific collocations obtained through KWICs, together with their frequency information provided in Tables 5.17 and 5.20 in Section 5.2.1, constitutes evidence in support of the existence of collocational features of CE, one of the identifiable features of ‘Chineseness’.

The findings in Section 5.2.1.1 reveal overused collocations in CE, which are strongly suggestive of CE users' preference. For example, the collocations *minority AREA*, *poverty/earthquake/disaster-stricken AREA*, *superior MAN* and *western PEOPLE*, from Table 5.6, are acceptable in BrE, as shown by the BNC search. If C. Gao (2009) is correct in saying that “acceptability is the benchmark for whether the expression is correct or not”, then these collocations (and others like them) clearly conform to the norms and are therefore correct English. Overused collocations provide statistical evidence of CE users' preference to use certain collocations.

Preferred collocations are very important in that they constitute features that could be described in sociocultural terms and, for that reason, they provide evidence of local norms that should be incorporated into the local English curriculum for teaching and learning. As Sharifian argues: “ELT pedagogy should involve creating natural opportunities for learners to engage in reflecting and explicating their cultural conceptualisations” (Sharifian, 2013, p.10). Recognition of variety-specific collocations thus constitutes strong evidence of local conceptualisations, which are especially important for English language teaching in that they should be considered for inclusion in the local English curriculum.

6.2 Word order in CE collocations

One of the most marked differences between CE and BrE is CE users' clear preference for pre-modifiers before nouns being modified. The phenomenon is referred to as word order, which indicates the linear sequence of word categories in noun phrases (i.e., modifier-modified word order in a noun phrase). As a matter of fact, Chinese is “one of those languages that rely heavily on word order as an underlying marking feature for meaning” (Ho, 1993, p. 138).

The recognition of this fact became evident as early as half a century ago, as summarised by Chao (1968) when he says ‘It is often said that all Chinese grammar is syntax, all Chinese syntax is word order, and therefore all Chinese grammar is word order’ (p. 260). Thus we can see the centrality of word order as part of an L1-conditioned cognitive disposition among CE users. There are a number of arguments which to some extent support the connection between word order in language and the mindset of the speaker. For example, Firbas (1992) emphasises that the word-order system of a language can be understood in a more comprehensive way if it is compared with that of another language, preferably typologically a distant one.

Based on Li and Thompson (1981), Lust and Chien (1984), Rutherford (1983),



Sun and Givón (1985), Tai (1973), Thompson (1978) and Tomlin (1986), the word order features of Chinese and English are summarised in Table 6.1.

Table 6.1: Comparison of Chinese and English word orders

Feature	Chinese	English
canonical word order	SVO/SOV	SVO
prominence	topic-prominent	subject-prominent
syntactic constraint	pragmatic word order	grammatical word order
principal branching direction	principally left-branching	principally right-branching

Although this summary may not seem directly relevant to the present study, it shows that, on a larger scale, in cases other than noun phrases, Chinese and English have different word order systems. Tai (1985, 1989, 1993) argues that a “modifier-before-head”(W. Jiang, 2009, p. 83)(or referred to earlier in this thesis as the ‘modifier-modified’ sequence) word order principle in Chinese is the universal principle governing Chinese word order as well as the grammatical rule governing Chinese linguistic structures.

Based on the findings in this thesis, there is evidence showing that the modifier-modified sequence before the head noun is preferred by CE users, resulting in the salience of sometimes lengthy pre-nominal modifying structures.

For example, in Figure 5.15, Section 5.2.2.2, 1,150 nouns (per million words) serve as modifiers of the 13 target lemmas, while the number is only 600 (per million words) in BNC. Quantitatively speaking, as a result, we could argue for the relative salience of pre-nominal modifying structures in CE. So word order is one of the most powerful devices used in Chinese to indicate subtle changes in meaning. This governing principle requires the word order of a modifier to appear before the noun it modifies in Chinese, no matter whether the modifier is an adjective, an attributive clause, a prepositional phrase, an infinitive verb phrase (e.g., 可愛的孩子, 媽媽買的衣服, 朝南的房子, 跳的高度) (Y. H. Liu, 2001, pp. 46-47). Most of the Chinese syllables are morphemic; any alteration to their sequence will therefore lead to significant change to their meaning.

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



(1)a 四川人**不**怕**辣**

Sìchuānrén bù pàlà;

Sichuan person not fear spicy;

Sichuaners do not fear (their food) being spicy.

(1)b 湖北人**辣****不**怕

Húběi rén là bùpà;

Hubei person spicy not fear;

(Their food) being spicy is not a fearful matter to Hubeiners.

(1)c 湖南人怕**不**辣

Húnán rén pà bùlà.

Hunan person fear not spicy.

Hunaners fear that (their food) is not spicy (enough).

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

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

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



(1)c Verb + negator + adjective (怕不辣, V-O structure, in which the object position is filled by a negated adjective or stative verb ‘(being) spicy’)

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

Generally speaking, the three phrases express very similar ideas; that is, Sichuaners, Hubeiners and Hunaners are alike, in that they all like spicy food very much. Word play apart, varying the word orders as a rhetorical device to convey subtle differences in meaning is not rare in Chinese. Word order (here in this thesis limited to noun phrases) is therefore an important aspect of the language disposition, influencing the structuring of information that CE users rely on heavily to express or manipulate meaning for special semantic effect. Rearrangement of the word order might not only change the meaning, but it could also shift the intended rhetorical effect of a given idea. For example, compare the effect of telling one’s teacher 我不完全懂 (wǒ bù wánquán dǒng, I did not understand fully [about what you said]) and 我完全不懂 (wǒ wánquán bùdǒng, I did not catch anything [you said]). While the five morpho-syllables (‘not completely V’ vs. ‘completely not V’) are identical, varying their order would not only change the scope of negation and result in



rather different meanings, but it would also result in rather different rhetorical effects. Thus in Mandarin, word order plays a crucial role; in that it helps people deliver information and ideas in subtle ways. To play safe, therefore, the CE user would be inclined to use the structure they feel more comfortable with, thus making the modifier-modified sequence in their noun phrases statistically such a distinctive feature in CE.

Thus the empirically supported answer to the second research question ‘Are there colligational preferences in CE?’ is: Yes, there are distinctive colligational features in CE. One of the features of CE found in the previous chapter is the approach to word order in CE, where users tend to demonstrate a clear preference for putting the modifier in front of the modified rather than after it, thus forming a modifier-before-head sequence when constructing collocations. More specifically, within a noun phrase, most modifiers (adjectives in this study) occur prior to the head nouns.

Furthermore, our analysis of CE data yielded far fewer cases of the *N + of* structure compared with BrE, which indicates a postmodifying or reverse word order in the structure of elements within a noun phrase. So a more specific answer to research question two ‘Are there colligational preferences in CE?’ is that colligational preference is indeed found in CE collocations, in particular



preferred word order patterns. According to Lian (2010, p. 25), Mandarin Chinese, like other Chinese ‘dialects’ or languages in the Sino-Tibetan family, is typologically an analytical language, with little inflectional morphology to convey grammatical relationships. An analytic language is marked by the relatively frequent use of function words, auxiliary verbs, and varying word order as principal means to express syntactic relations, rather than relying on inflected word forms. As a result, free morphemes, which are often separate words, are used very commonly in grammatical constructions along with word order. In sum, one of the distinctive features of the L1-conditioned cognitive disposition of CE users is derived from the fact that Chinese languages and dialects have relatively restrictive word orders, often relying on the order of constituents to convey important grammatical information. In contrast to this, inner-circle varieties such as British English can convey grammatical information through inflection, which allows for more flexibility in terms of word order (e.g., both premodifying and postmodifying structures in a noun phrase). In the case of CE, the modifier-modified sequence is a word order structure which is used with higher frequency than can be accounted for by chance, as reflected in the preference of CE users in the CEC corpus. In the 13 target lemmas in Figure 5.3, *WAY* and *FACT* were taken as example because their frequencies of occurrence are close in CEC and BNC, and not very topic-specific. Following the patterns of Tables 5.21 and 5.22, I searched the



pre-modifiers (span=L-1) and post-modifiers (span=R-1) of *WAY*, *FACT* and *THING*, and found that, statistically, there is a clear tendency that the modifier-modified sequence is preferred (Figure 6.4; compare corresponding BNC numbers in Figure 6.5).

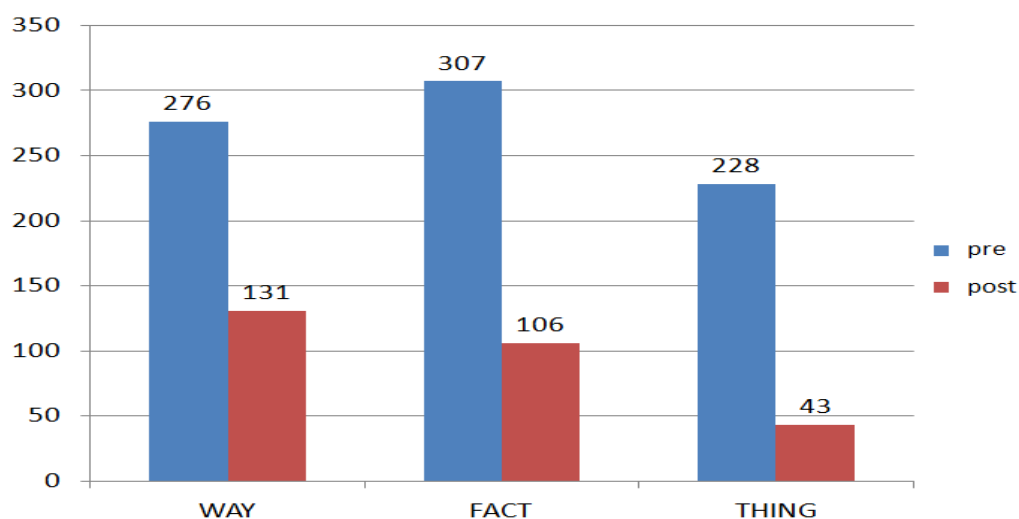


Figure 6.4: Contrasting the numbers of pre-modifiers and post-modifiers of *WAY*, *FACT* and *THING* in CEC (per million words)

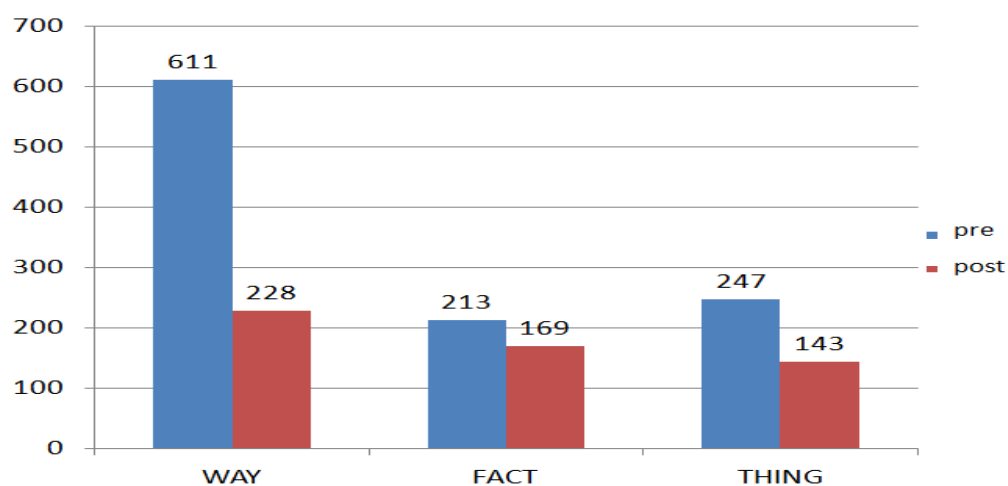


Figure 6.5: Contrasting the numbers of pre-modifiers and post-modifiers of *WAY*, *FACT* and *THING* in BNC (per million words)

As indicated in Chapter 2, corpus linguistics views the meaning potential of language probabilistically, which makes differential frequency information (as in Figures 6.4 and 6.5) so significant. *WAY*, *FACT* and *THING* are samples for our study in this part, because we need high-frequency words that are not topic-specific and are used by general English users. Table 6.2 shows the ratios of pre-modifiers to post-modifiers in CEC and BNC of these three lemmas:

Table 6.2: Summary of ratios of pre-modifiers to post-modifiers of *WAY*, *FACT*, *THING* in CEC and BNC

Ratio of pre: post	CEC	BNC
WAY	2.1:1	2.7:1
FACT	2.9:1	1.3:1
THING	5.3:1	1.7:1

While the ratios in Table 6.2 are indicative of a general tendency or preference for pre-modification in CE noun phrases, more in-depth research is needed before such a trend could be confirmed.

To sum up, CE users tend to demonstrate a tendency or preference for putting the modifier in front of the modified, thus forming a tendency of modifier-before-head sequence when constructing collocations, which is why it

is such a salient colligational feature in CE. To make the point even clearer, Chinese languages or dialects rely on word order to construct sentences and meanings. It is generally true that a change in word order would trigger a change in the meaning of the sentence. By virtue of the word order patterns in Chinese (the patterns are restricted to the collocations under investigation in this thesis, while word order beyond a noun phrase in a sentence is beyond the scope of our investigation), our primary focus is on how morpho-syllables are sequenced or structured to produce meanings, and how word-order based meanings articulate with the meanings contributed by individual, constituent morpho-syllables themselves. One may argue that this is to some extent true of any language, but it is very markedly so in the case of Chinese (varieties) compared to English. The reliance on word order to make meanings inevitably constitutes a preferred mode of expression in CE, which helps to explain the left-branching tendency, as well as the preponderance of the modifier-modified structure in the CE noun phrase.

Historically speaking, language contact may also have contributed to the salience of this modifier-modified preference in noun-phrase collocations in CE. Sustained language contact between Chinese and English initiated since the *Opium Wars* (1839-42; 1856-60) allowed for mutual influence between these two languages. And the “Vernacular Movement” (*báihuàwén yùndòng*), which



sprang up after the *May Fourth Movement* in 1919, intensified the Europeanisation of Chinese. Research shows that after the *May Fourth Movement*, many writers (e.g. Lu Sun, Yu Dafu etc., during *the New Literature movement* from 1910s to 1920s) preferred the modifier-modified structure, sometimes leading to very long premodifying elements before head nouns. The following example, adapted from the vernacular literature of that period, is particularly instructive, showing a long sequence of premodifiers (underlined) before the head noun (as bolded).

1. 秦國本來是專講究武器，年年不斷地招募新兵，看百姓不值一錢，只將百姓的辛苦勞力全部壓榨出來，只用到打仗殺人等事情上去的一個**國家**。

(郁達夫《故事》) (Cited from Y. Xie, 1990, pp. 84-85)

English version: ‘Qin is such a **nation** which was meticulous with weaponry, recruited new soldiers every year, saw commoners as worthless people, exploited their sweaty labour to the full, and used all this for fighting and killing in battles and the like.’

As shown in this contrastive example, to render (1) idiomatically into English, all the pre-modifying clauses in Chinese must be packaged as post-modifying clauses in English.

A similar trend could be observed in written vernacular Chinese (*báihuàwén*)



since the *May Fourth movement*, which has been criticised by linguists as one feature of adverse Europeanisation of the Chinese language (e.g., G. Yu, 1987; Y. Xie, 1990). However, after a hundred years' mutual influence, the effects of Europeanisation of Chinese may have become deeply entrenched in the way speakers/writers of Chinese use their first language with regard to their preference for the modifier-modified order for noun phrases. More importantly, it is highly plausible that such a preference is transferred to their English usage, resulting in a salient feature of CE in its development toward an autonomous variety of English.

6.3 Rhetorical features

The third research question “Are there semantic and prosodic preferences” in CE asks whether there are any distinctive rhetorical preferences in the collocations used by Chinese speakers of English. By ‘rhetorical’, I mean “the ways of using language effectively” (Matthews, 2006).

The hypothesis was that preferred rhetorical devices in terms of semantic preference and prosody may constitute a unique feature in CE. In light of the findings in Chapter 5, especially descriptions of *PEOPLE*, *DEVELOPMENT*, *WAY*, *CHILD*, *SPIRIT* in Section 5.2.3.3 and Section 5.2.4.3, we may now



conclude that the answer to this question is a positive one, with two issues arising from this finding: a) diction; and b) formation of collocations.

By ‘diction’, I mean the words and patterns which people choose (not) to use. This point has been presented and illustrated with the help of frequency information in Chapter 5. For example, the high-frequency use of a modifier adjective *all-round* (cf. Table 5.19; see also Figure 5.10), or the high-frequency usage of collocations such as *western people*, *foreign people*, *common people*, which all have alternative ways of saying the same thing (e.g., *people from a western country*, *people from a foreign country*, and *ordinary people*, respectively). All this constitutes evidence in support of the existence of preferred diction among CE users. Examples of variation between CE and BrE suggest that the choices of diction made by Chinese speakers are motivated rather than arbitrary.

One possible motivating factor, as proposed here, is the Chinese-specific cultural nuances of these items. The statistical salience and use of such terms are arguably driven by the local conceptualisations and cultural realities embedded in the Chinese equivalents of these items. For example, the word *guānxi* (meaning ‘relation’, ‘relationship’, ‘network’ or ‘backdoor practice’) (Xu, 2010, p. 33) reflects a Chinese “cultural schema” (Sharifian, 2013, p. 7);



moon cake, *lucky money* reflect a “cultural category” of CE users (ibid. p. 8); while *golden rice bowl* (a secure high-paying job) is an example of “cultural metaphor” (ibid. p. 8). In these examples, we can see that there is no alternative English expression, in BrE or any other inner-circle variety, which could replace the “cultural conceptualisations” (Sharifian, 2003, 2008, 2011) in the Chinese expression. The absence of the exact equivalent of the concept may eventually account for their preference of diction in the English they used, based on shared cultural conceptualisations (e.g., *common PEOPLE* rather than *ordinary PEOPLE*, since the former is believed to be closer to the concept of *rénmín*). This phenomenon is observable in Mandarin too: People from the Cantonese speaking region could be identified (when they are speaking *Putonghua*) by their diction of saying ‘見到’ (jiàndào) rather than ‘看到’ (kàndào) in sentence like ‘不要給我見到’ (meaning: Don’t show this to me). People using Putonghua as their L1 would be saying ‘不要給我看到’. Though 看 and 見 are synonyms, the choice of diction in ways how they collocate with 到 gives some clue about the speakers’ L1 background. So an individual’s repertoire of ‘choosing’ words and their collocates in their first language arguably influences or informs the way they select words in English, resulting in the development of patterned features associated with the variety of English they are using (in our case, CE).



While deviation in diction concerns the words used or not used (resulting in overuse and underuse), the formation of collocations can be seen as the manner in which people ‘combine’ language elements, resulting in distinctive patterns. This study explores these patterns more deeply, namely in the semantic preference and prosodic features that tend to be neglected in WE research.

The findings in Sections 5.2.3.3 and 5.2.4.3 reveal that the core semantic preferences of these CE items are similar to those in BrE. However, deviations have been found with regard to some items that bear cultural significance in Chinese. For example, in Tables 5.29 and 5.30, we have argued that main verbs that collocate with the word DEVELOPMENT in CE tend to be more positive compared with its counterpart in BrE.

In Figure 5.5, we have shown a higher ratio of using the collocation *common PEOPLE* in CEC than BNC, suggesting that this collocation has special connotation among CE users, probably resulting from a strong motivation of transferring the collocation in Chinese (普通人 pǔtōngrén), rather than its synonym *ordinary PEOPLE*. In Table 5.34, we have shown that those adjectives which collocate with the word *SPIRIT* are perceived as having a “good prosody” (Louw, 1993, p. 157). To illustrate, we will now examine more closely the semantic and prosodic preference of the unique collocation



innovative SPIRIT, whose KWIC is presented in Figure 6.6 below.

Among the concordance lines, we can find that *innovative SPIRIT* is used in a variety of fairly general registers and thus not specific to only political slogans. For instance, the domains in which it occurs vary, from military (*PLA*) and music (*flute*) to education (*former Peking University President CaiYuanpei*). Subjects that can have *innovative SPIRIT* can be individuals (*student, scholar, technicians*) or collective groups (*Chinese society, literary circles, faculty*).

In a word, this collocation is used in general English and covers a wider range of topics and registers within its scope. One key point is that, the collocation *innovative SPIRIT* is not used in BNC. So this collocation arguably constitutes evidence in support of both research questions one and three.

Taken together, the results presented under the semantic and prosodic patterns of the target words suggest that while CE users tend to adhere to an inner circle norm, they have no reservation forming distinctive patterns in distinctive ways in some cases, typically when Mainland Chinese sociocultural realities are foregrounded. For those lemmas like *DEVELOPMENT, PEOPLE, SPIRIT*, we can see preferences in the selection of main verbs and adjectives, which are distinct from their counterparts in inner circle varieties.

KWIC

e with the times, work hard with a pioneering and innovative spirit, unswervingly take the road of Chinese
 mote a patriotic national spirit, a reformist and innovative spirit of the times, and a collectivist and so
 academic titles, well-balanced age structure, and innovative spirit. There are 32 full-time teachers, includ
 gh level educations, reasonable age structure and innovative spirits. The present faculty includes 20 Ph.D.
 difficulties and hardships in an indomitable and innovative spirit and made important advances in building
 loped morally, intellectually and physically with innovative spirit, who have the knowledge of finance theo
 ke it as our basic task to cultivate talents with innovative spirits. We will encourage independent innovat
 Practice of Education and Cultivating People with Innovative Spirit, to discuss new approaches to education
 is a large faculty with excellent traditions and innovative spirits. It has been fostering highly qualifie
 r Ph.D students to give full expression to their innovative spirit. In addition, a series of activities suc
 e platform and core technology. Her knowledge and innovative spirit in business management helped her compa
 f bureaucracy, communication problem, decrease of innovative spirit, and the organization tends to fossilize
 se projects. A new generation of scholars full of innovative spirit have become the backbone of the univers
 nt while upgrading those in active service. In an innovative spirit, the PLA endeavors to improve its train
 pment. Aiming at nurturing talented students with innovative spirit, Tsinghua has developed a complete syste
 f-management and to cultivate the team spirit and innovative spirit. Nowadays, team spirit has been a remar
 e technicians who really know techniques and have innovative spirit, especially senior technicians grasping
 n education. Chinese society needs the wisdom and innovative spirit shown in this campaign. This is a farsie
 ion. The PLA develops its military theories in an innovative spirit, and explores the laws of building the
 hought of Three Represents, worked together in an innovative spirit, and made major achievements worth celel
 's beliefs in traditional values as well as their innovative spirit. The cucurbit flute is a musical instrum
 arts design students with advanced consciousness, innovative spirits and highly sense of social responsibil
 n fresh air into literary circles with an unusual innovative spirit. A major change in these periodicals is
 ormer PKU president Cai Yuanpei, demonstrating an innovative spirit. The core idea of Yuanpei is to demolish

Figure 6.6: Sample KWIC of *innovative spirit* in CEC



After connecting findings with the linguistic and rhetorical background of CE users, I believe the findings of the previous chapter offer a positive answer to the third research question. Prevailing preferences in collocation, colligation, semantic preference and semantic prosody prove that there is an element of subjectivity in the patterns, which would seem to suggest a shared lifeworld within broadly defined Mainland Chinese cultures. So the mode of thinking or conceptualising of a more or less homogeneous ethnolinguistic group, through which its language acquires colour and character, is at work from the outset. The motivation for CE collocational features is the inner character which dwells in the language like a soul. Since lifeworlds of all people are equal, it is not fair to elevate one and look down upon another.

6.4 Collocational features of CE: A theoretical framework

The final research question asks whether it is viable to propose a theoretical framework for researching collocational features of China English based on the outcome of this investigation. We saw in Section 2.5.2 that frequency information of words from a corpus does appear to give useful information about the possibility of encountering them in written and spoken communication, suggesting that information of this kind may enable us to pick out sets of collocations which eventually may help us to codify CE collocational features to inform WE and ELF research. Section 5.2 showed that collocational patterns of four kinds (i.e., collocation, colligation, semantic preference, and semantic prosody) revealed variations which are



characteristically different from those of inner circle varieties quantitatively. Those results presented in Tables 5.6 and 5.17; Figures 5.12 and 5.14; and Tables 5.29 and 5.30 along with their analyses in the relevant Sections constitute supportive evidence of the existence of the four kinds of patterns mentioned above. They suggest that these features are different from previous findings of CE-feature research, which tended to rely on specific words as evidence. Our findings in this study have uncovered some covert features hidden behind or underneath such overt features. For example, we may easily tackle the collocation ‘plain MAN and WOMAN’ (cf. Section 5.2.1.2), ‘innovative SPIRIT’ (cf. Section 6.3), but in regard to target words like DEVELOPMENT and SPIRIT, it may not be apparent to other varieties of English that certain types of words would tend to be attracted/co-occurred with them in CE. As such, what are distinctive to SPIRIT and DEVELOPMENT are not the specific collocates, but more subtle ‘colour’ of all the collocates. I hope to have demonstrated that such covert features are at least as important and interesting in linguistic research, and that they deserve closer attention and further research.

The purpose of proposing a theoretical framework for researching CE as an emerging variety of English is to try to reveal the lingua-cultural categories that may constitute CE collocational features. Furthermore, it is hoped that the framework proposed in this study can help us visualise the inter-relationship between overt and covert features of the target words being examined, thereby helping us to gain a broader view of CE as an emerging new variety of English.



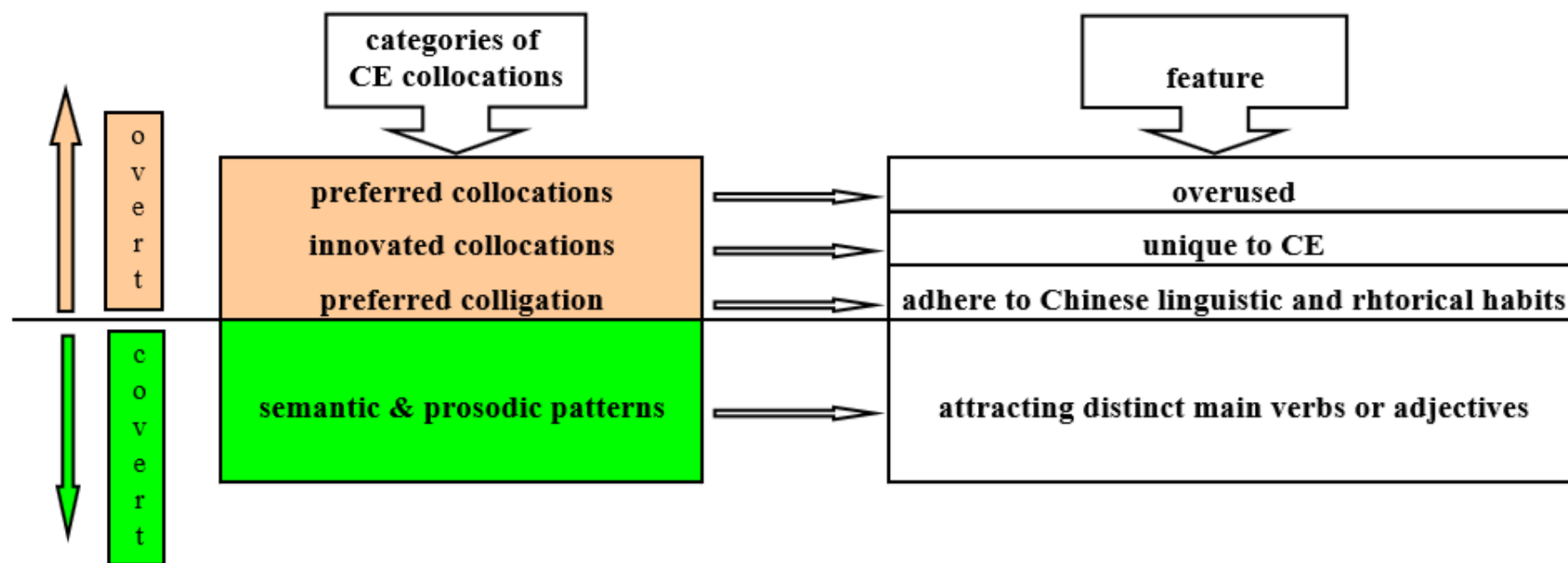


Figure 6.7: A proposed framework for analysing collocational features in CE

More fundamentally, we need to emphasise the explicitness as well as the implicitness of CE in terms of collocational patterns. Features that serve as indicators of each of the “collocational categories” (Sinclair, 2004, p. 174) of CE should also be pointed out. So with this purpose in mind, in the following part, an attempt will be made to outline a framework for analysing collocational features in CE based on the findings and evidence unveiled to date (see Figure 6.7).

This framework has the merit of describing and summarising the findings in Chapter 5, and at the same time clarifying how we could incorporate all the aspects involved into the instrument. The primary focus is to enable us to make a clear distinction between overt and covert features in CE collocations.

The framework consists of two levels, labelled ‘overt feature’ and ‘covert feature’ respectively. At the overt level, categories of CE collocations such as preferred collocations (e.g., *common PEOPLE*), innovative collocations (e.g., *in an all-round WAY*), preferred colligation (e.g., *N’s + N of*) are indicated. At the covert level, the CE semantic and prosodic features (e.g., Table 5.29, 5.31) are presented in accordance with the words they attract to produce the specific semantic prosody.

This framework also gives some indication of the operational dimension, showing how further studies on CE collocations may be carried out. The column on the right with the name ‘feature’ refers to the targeted findings that could be observed and pieced together after necessary processes of investigations are taken. These specific research components for observations would help reveal their corresponding features listed in the middle column named ‘categories of CE features’. For example, overused collocations may reveal that CE users have ‘preferred collocations’. When we find ‘collocations unique to CE’, we may also find ‘innovative collocations’, etc. These two columns would situate ‘results that have been found’ and ‘the categories they belong to’ in a conceptualised way.

6.5 Summary of chapter

This chapter sums up proposed answers to the research questions set out in chapter 2, and presents interpretations of the findings obtained in Chapter 5 with the help of more qualitative findings obtained from sample concordances. The findings allow us to construct a theoretical framework for uncovering variety-specific features beyond those statistically salient, overt features. The significance of this study in terms of its contribution to its field will be discussed in the next and final chapter.



CHAPTER 7

CONCLUSION AND IMPLICATIONS

This thesis has attempted to address several research questions regarding the China English (CE) features in terms of co-selection and collocational patterns. In this final chapter, I will try to tease out the significant contributions of the findings in this study to research on World Englishes (WE) and English as a lingua franca (ELF), and draw implications for further investigation in CE.

7.1 Recapitulation

This study set out to uncover the collocational patterns that are distinct to CE users in comparison with inner circle English norms, by using quantitative methodologies which take advantage of word-frequency information. It is hoped that the methodological design, analytical procedures and empirical findings in this study will go some way to helping us better understand the nature of CE in future studies. As indicated by the title of the thesis, primary attention has been focused mainly on questions concerning collocational features of CE.



The empirical findings as reported and discussed in Chapters 5 and 6 point to a number of characteristics pertaining to CE as an emerging variety of English:

- a) By and large, CE shares many of the core collocational patterns in BrE, as shown in the similar use of high-frequency words (see Sections 5.2.3.2 and 5.2.4.2).
- b) However, CE does possess her own preferred collocational patterns which deviate from BrE usages. Such CE-specific usages occurred not only in directly observable, concrete lexical co-selections (collocations), but also in groups of more abstract grammatical co-selections (colligations), as well as semantically and attitudinally preferred co-selections (semantic preference and semantic prosody) (see Sections 5.2.1 to 5.2.4).
- c) The four parameters (collocation, colligation, semantic preference, and semantic prosody) may serve as discovery or analytical procedures for uncovering further linguistic features in CE, with CE-specific collocations and colligations constituting evidence of their semantic preference and semantic prosody.
- d) The CE-specific collocational features identified in this study broadly reflect, and are in harmony with typical, culturally conditioned behaviours of CE users, consciously (in which case, intentional) or subconsciously (in which case, attitudinal). In either case, CE-users' English outputs are

guided by such lingua-culturally informed intentions and/or attitudes.

- e) Norms of CE, which may be broadly characterised as an amalgamation of a lexico-grammatical common core in Standard English (here, BrE) plus CE-specific collocations, are correctly understood by and widely shared among fellow CE-users.
- f) CE users tend to overuse words of high semantic generality, which serves as a way to fill lexical gaps where necessary and appropriate (e.g., more general/common adjectives such as *all-round* are used to modify nouns).
- g) CE users tend to prefer explicitness in collocation despite semantic redundancy (e.g., *true fact*).
- h) CE collocations are systematic, consistent, and predictable in nature.

It has been argued in this thesis that CE users' English can be more systematically accounted for if we are aware of the kinds of patterns and interrelations which words contract with one another. We should note that it is the patterns created by the more frequent core words in the lexicon—words such as *DEVELOPMENT*, *SPIRIT*, *CITY*, *CHILD*—which are the most problematic to describe as well as to teach and to codify lexicographically. This study is in line with Carter's (2012) point of view in regard to vocabulary use:

It is commonly assumed that using words entails a creative development of the resources of the language, particularly in the *selection* of items from our lexical store; but many lexical items are either themselves patterns or form part of patterns which are quite fixed and stable and which are used routinely in relatively predictable situations. Meanings can, of course, be uniquely generated but stability is a pervasive feature of normal vocabulary use and it is clear that there are numerous communicative contexts in which language can be used formulaically. (p. 84, italics in original)

In Chapter 6, I suggested that when we look at CE from the same point of view, we may also find *clines* of fixity in patterns of usage. It is interesting that collocational patterns of these types do not figure in EA categories because they are not errors. The basic assumption is that, however important it is to account for and understand lexical patterns of inner circle varieties of English, it is also valuable to identify and codify features of collocational patterns of expanding circle varieties of English as English belongs to all its speakers while communication involves all people around the world. Only after we recognise equality in using English can we eliminate linguistic imperialism (Phillipson, 1992) and language bias or prejudice, neither of which will do any good to communication between people across borders.



It is my belief that corpus-based analysis can help unlock many more collocational patterns that are characteristic of CE, and that more research along this line will enable us to define CE in more concrete terms lexico-grammatically. In my view, the future of CE looks promising in this regard.

7.2 Implications

Some implications may be drawn from this study. Firstly, as an emerging variety of English, CE clearly differs from other varieties of English in distinctive ways. If CE allows its users to communicate their meanings more or less clearly, it is largely because CE shares a lot of the Standard English features. So, if CE is different, it differs not only in *what* its users can express, but *how* they do it. As a priority in further research, therefore, there is a need to build on the common core that CE shares with Standard English by laying bare its lingua-culturally unique content words and collocational patterns. In this regard, I hope to have demonstrated that corpus-based analysis as exemplified in this study holds the key to this research agenda.

Secondly, sections that report overused and innovative patterns in CE (Section 5.2.1, pp. 171-209) have the potential to tell as something about the values of



society and context they come from (here in this thesis, the Chinese context). The identification of overused patterns carries considerable significance which goes well beyond a ‘bean-counting’ or mere tallying activity, for “[f]requency can be an indicator of markedness” and a research tool for “unpacking ideological assumptions” (Baker, 2010, p.125). Thus, from this point of view, our findings, derived from frequency data of recurrent collocations and manifested as overused/innovative patterns in CE, clearly have reference value in the revision of ELT curricula in Mainland China. Inclusion of distinct CE patterns of collocations in the syllabi and textbooks are arguably rational and appropriate since they represent social values that are important and instrumental for indexing CE identity. The main goal of this thesis is to uncover recurrent CE collocations, leaving the task of providing a longer and fuller list of CE collocatons for future research.

Thirdly, such findings may also contribute to enhancing the quality of intercultural communication involving CE users. As has been discussed in Chapters 1 and 2, two major factors constitutue the current context of CE users’ use of English: a) since 1980s, English has been taught in Chinese via education from primary to tertiary levels throughout the country as a compulsory subject, with ‘being able to communicate’ with people from other cultures as the target learning outcome; b) ICT-related technologies via the internet has largely



‘softened’ cross-border communication, which for about a decade keeps generating (typically electronic) platforms and multiplying opportunities for ELF communication, possibly including Chinese-Chinese dyads. It is against this background that researching CE collocational features is seen as conducive to enhancing mutual understanding in ever-expanding intercultural communication.

Another implication for finding collocational features of CE lies in the possibility that it could help establish CE as one member in a ‘pluricentric’ (Kloss, 1978; Clyne, 1992; Clyne & Sharifian, 2008) family of Englishes in the long run. Because CE should provide “a national variety with at least some of its own (codified) norms” (Kloss, 1978, p. 66) before she could be legitimated as one player in interactional activities in English use. The necessity of being inclusive in the pluricentric circles of Englishes is sociolinguistically far-reaching, for

[P]luricentric languages are both unifiers and dividers of peoples. They unify people through the use of the language and separate them through the development of national norms and indices and linguistic variables with which the speakers identify. (Clyne, 1992, p. 1)



CE collocational features help mark group boundaries which are indications of who belongs and who does not. Potential language conflicts may also occur if we stick rigidly to inner circle variety norms in English use and ignore the high-frequency/overt features, as well as the covert features in English teaching.

Take the term *tofu* (*bean curd*, 豆腐) for example. The ways to express the Chinese 豆腐 in English (and thus how to teach students) has undergone several changes as manifested in Chinese-English Dictionaries since the 1960's, where people are supposed to find answers regarding how best to express ideas and things. Half a century ago, someone trying to know what 豆腐 was in English would find the following entry in Mathews' dictionary (1960):

豆腐: bean-curd, a preparation which, in its liquid state, is almost identical with cow's milk as regards its nutritive constituents. (*Mathew's Chinese-English dictionary*, 1960, p. 939)

According to the norm set by the compiler of the dictionary, there should be a hyphen between *bean* and *curd* (i.e., *bean-curd*). However, we found that the hyphen was removed somehow later, for example, in a Chinese-English dictionary in the 1980s which said:



豆腐: bean curd. (Ding, 1985, p. 243)

Some twenty years later, however, when one looks up this word in a more recent Chinese-English dictionary, one would find something like:

豆腐: beancurd, tofu. (Yuan & Church, 2006, p. 31)

Notice that an alternative for *beancurd-tofu*—was provided. Today, it is really more common practice to use only the word *tofu*, as we know. A shift has been made, therefore, in the most idiomatic translation of 豆腐 (*bean-curd*, *bean curd*, *beancurd*, *tofu*) through the years. This example shows that as time goes by, norms of expressing culture-specific terms in English could shift from an inner circle rendition to one in CE—a pluricentric movement and lexical variation across time. In a word, one implication of research for CE collocational features would be to foster the legitimation and acceptance of other CE-specific words among its users, in China and beyond.

7.3 Significance and values

In the current highly multilingual context, English is not only being used as a contact language in diplomacy, international trade and business, and tourism,



but also a communication instrument that is mostly employed by individuals from different language backgrounds in their daily communication with others. It is “the world’s default mode of communication” (McArthur, 2002, p. 13). British English used to be the most influential variety of English, but today it has been replaced by American English due to a shift in “political power and the resultant diffusion of American culture and media, technological advance and the rapid development of communications technology”(Kirkpatrick, 2007, p. 55).

However, it may not be always so. “With China’s emergence as a world power, with its increasing integration into the world system, China will need English to project its own presence on the regional and the international scene” (Bolton & Tong, 2002, p. 180). CE is likely to become a “future power” (Xu, 2010, p. 282) since “with every passing day an ever-increasing number of Chinese speakers of English are added to the multi-million community of Chinese-English bilinguals in the Mainland” (ibid.). Indeed, CE may one day “be in the forefront of the development of English language” (Deterding, 2006, p. 195). It thus seems possible to predict that CE may exert considerable influence on the further development of the English language worldwide. Thus one of the fundamental issues to be addressed is “What are some of the salient linguistic features of China English?” (He & Li, 2009, p. 70).



Essentially, the challenge for researchers in ‘China English’ is:

1. to identify and select salient linguistic features of ‘China English’ which are widely represented and attested in different regions on China;
2. to come up with a rigorous codification scheme for those linguistic features selected; and
3. to convince the national education authorities that the ideal teaching model of English in China should be based on NS-based norms and standards, to be supplemented judiciously with selected features (or variants) of China English (He & Li, 2009, p. 86).

Research into CE will contribute to its development as an emerging variety in the WE family. Furthermore, such research would be extremely useful for English education in China, since English plays an increasingly important role in educational circles, and in meeting people’s communication needs, both of which have been discussed at length in Chapter 2.

The fact that China has been developing fast for over three decades since the early 1980s makes it self-evident that more and more Chinese people are in contact with the rest of the world than ever, and this momentum is expected to continue in the near future. Despite the fact that the first known use of English



in China is dated as far back as 1637 (Chapter 3, Bolton, 2003), there was little awareness of this, and little research attention to it, until a few decades ago. This may be due to the fact that China belongs to the ‘expanding circle’ of World Englishes (Kachru, 1985) and was not traditionally considered as a variety; typically reference would only be made to the ‘inner circle’ for norms and standards.

However, researchers increasingly argue that “native Englishes, indigenised Englishes and English pidgins and creoles have all developed by the same kind of natural restructuring processes” (Mufwene, 2001, p. 113). This point is echoed and summarised by Kirkpatrick (2007) as follows:

All varieties of English have been influenced by contact with other languages and have adopted vocabulary items from local languages (...). All varieties must, on the one hand, reflect the cultural realities of their speakers and, on the other, be adapted enough to allow international communication. (Kirkpatrick, 2007, p. 34)

Now that more evidence is available, it seems sensible to talk about CE as a developing variety, for there are hundreds of millions of people using it now and even more will do so in future (Xu, 2010).



The careful description and judicious codification of CE may allow us to talk about a Chinese variety of English someday, and to validate CE as a new member of the WE family, due to the simple fact that Chinese learners and users of English constitute the largest group of bilingual users of English in the world. The language they use, the Chinese variety of English, may become a widespread, systematic, rule-governed variety among Chinese speakers. Codifying CE features can advance the course of World Englishes in the context of China.

In 1965, Fishman wrote an article in *La Linguistique* with the title ‘Who speaks what language to whom and when’, which was often quoted as a classic, succinct definition for sociolinguistics. Language is an indispensable semiotic resource for meaning-making in social, interpersonal interaction. “A sociolinguistic perspective requires exploring how language is used to establish a social context while simultaneously exploring how the social context influences language use and the communication of meaning” (Bloome & Green, 2002, p. 396). Sociolinguists are therefore often interested in analysing how the identity of a person in a social group relates to the way they use language. They investigate issues such as the linguistic similarities and differences between certain groups of people, and the ways in which social variables such as geographical region and nationality background impact on their language use.



In the field of language learning, there has always been interest in language and identity. Theorists who have been influential include C. West (1992), Norton (1997), Pennycook (2001), to name but a few, among many others. They argue convincingly that the use of language(s) in social interaction is strongly associated with the desire for recognition, affiliation and respect from others. Thus when it comes to the issue of identity, learners of a given language share similar rights with those who use it as their first language

[I]f English belongs to the people who speak it, whether native or nonnative, whether ESL or EFL, whether standard or nonstandard, then the expansion of English in this era of rapid globalization may possibly be for the better rather than for the worse.” (Norton, 1997, p. 427)

Research on the intricate relationship between language and identity may be traced back to the influential work by Le Page and Tabouret-Keller (1985), who showed that every speech act an individual performs is an ‘act of identity’. Personal use of language reveals their sense of social and ethnic solidarity or difference. At the same time, people also tend to have powerful (if unconscious) stereotypes about the norms and standards of their own language and those of others—often at variance with observable behaviour.



Patterns of language use play a crucial role in establishing relationships within regional and social communities. The mere act of naming a variety spoken by members of a community, one that was previously subsumed under some (more prestigious, albeit incorrect) version of another variety, tends to confer a particular status to that variety and possibly its users. To describe it would enhance its status even further. According to such ‘linguistic nationalism’ (Le Page & Tabouret-Keller, 1985, p. 234), there is common progression in the relationship between isolating and naming a group of people and the language they use:

1. A group of people is named in accordance with one of these criteria:
geographical provenance (e.g., ‘the Americans’); common parentages (e.g., ‘the tribe of Judah’); common physical characteristics (e.g., ‘the Blacks’); common traditions (e.g., ‘the Christians’);
2. The group’s linguistic behaviour is denoted by an adjective form from such a term, together with the noun for ‘language’ (e.g., Bahasa Malaysia);
3. These adjectives become used as nouns; and the nouns not only denote the linguistic system felt to be the property of the group, language as used by them, but also connote the social values attached to the group. The name of the language is inseparable from the sense of being the usage of a group of speakers;



4. The system becomes detachable from the group, and comes to have some degree of autonomy as a linguistic system; the social values which are attached to the group now tend to be transferred to the system;
5. Once the system has achieved some degree of autonomy in people's thinking, it can be both reified and totemised (usually involving some body of doctrine, e.g., grammar, lexicon, by means of which the system can be taught) (Le Page & Tabouret-Keller, 1985, pp. 235-236).

Inspired by these views, I believe that concerted efforts to unravel patterns of CE would help determine the progressive stages of its evolvment and eventually help establish CE identity as a primary concern behind and main motivation for its use among CE users collectively. As can be observed in daily life, some forms of language are taken up (e.g., the expression *long time no see*, the words *dim sum*, *tofu*, *wanton*), either consciously or otherwise, while others are discarded (e.g., English slogans known almost to everyone in the 1960s and 1970s: *long live Chairman Mao*; *good good study, day day up*; *down with so and so*). This phenomenon represents a research topic in itself, and the identification and analysis of CE features lies within the scope of such a task. It can therefore be seen that research into CE features virtually amounts to an attempt to better understand the changing socio-cultural values and experiences of the largest ethnolinguistic group of English users in the world.



Since corpus-based methodologies and sociolinguistic approaches entail the collection and analysis of naturally-occurring language data, and place a great deal of emphasis on language-in-use, they share a number of fundamental tenets in common. It is therefore hoped that this study has exemplified and demonstrated the fruitful use of corpus research techniques in the sociolinguistic analysis of language variation, thereby contributing to its further development.

Finally, the present research aspires to be a significant contribution to ELT in China, given that the main concerns in second language learning are ‘*Who learns how much of what languages under what conditions?*’ (Spolsky, 1989). In this case, the ‘who’ and ‘what language’ are clear, but the ‘how much’ and ‘in what conditions’ require serious deliberation. Toward the goal of making CE an accepted, autonomous variety in the near future thanks to its established features and norms which are distinctive from inner circle varieties, the ‘how much’, together with the preferred pedagogic model of English in the ELT curriculum in China, need to be addressed.

Nowadays bilingualism is seen as “a normal, rather than a special condition” (Graddol, 2006, p. 117), setting attainable and appropriate English language targets (e.g., norms and acceptability of certain usage), “focus on the local



functionalities (i.e., local needs, functions and purposes) in the ELT materials” (Xu, 2013, p.1) is absolutely crucial for both teachers and learners in China. From this perspective, it is hoped that the present research may at some point contribute to shifting the paradigm of English language teaching and research in China towards finding a legitimate place for those statistically sound and linguistically salient CE features.

To conclude, while English is not the first language of the majority of Chinese people in Mainland China, it is learned by and used among speakers of different Chinese varieties throughout the country, and in other activities where they are engaged in communication with other users in the rest of the world to whom English may or may not be their first language. Given the tremendous linguistic diversity and dialectal variation across different parts of Mainland China, it is conceivable that CE is unlikely to be a single monolithic variety. The premise of this thesis is that a better understanding of the exact nature of CE, as it is used today, can be gained by a detailed and empirically-grounded analysis of its distinctive linguistic features.

From a broader perspective, since the new millennium the enhancement of China’s national economic strengths worldwide has gradually improved her soft power globally, allowing China 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, fostering the hope of achieving recognition and respect for the nation. In this global context, it is hoped that research into the distinctive features of CE will contribute to a better understanding of CE users as they are engaged in communication in English, both intra- and internationally.

It is the belief of the present author that a corpus data-driven approach towards better understanding what China English consists in should be promoted, and that CE has its rightful place as a significant member in the World Englishes family, especially when we consider the developmental cycles of new varieties of English proposed by Schneider (2003) as summarised in Table 7.1:

Table 7.1: Developmental cycle of new varieties of English (Schneider 2003, 2007)

Phase	Description
1. Foundation	English begins to be used in this country
2. Exonormative stabilisation	the variety is modelled on the imported one
3. Nativisation	reconstructing, establishment of new identity
4. Endonormative stabilisation	local norm gradually accepted
5. Differentiation	new variety, reflects local identity and culture

While Schneider's (2003, 2007) theory of developmental cycle deals with the diachronic question, how CE evolves into a new variety, Butler's (1997) theory of 'five criteria for an emerging variety of English' is also useful for addressing a synchronic concern, whether and when to consider CE as a new variety:

- (1) a standard and recognisable pronunciation handed down from one generation to another;
- (2) particular words and phrases which spring up usually to express key features of the physical and social environment and which are regarded as peculiar to the variety;
- (3) a history—a sense that this variety of English is the way it is because of the history of the language community;
- (4) a literature written without apology in that variety of English;
- (5) reference works—dictionaries and style guides—which show that people in that language community look to themselves, not some outside authority, to decide what is right and wrong in terms of how they speak and write their English. (Butler, 1997, p. 10)

As evidence gradually accrues, I think we may begin to think about considering CE as a new variety in the near future. The findings in this study provide supporting evidence for this. As described in Chapter 2, the debate on CE along



with its attendant research agenda has been one heated area among Mainland Chinese linguists at home and beyond. The overall stance of the local people towards the usage of English typical of CE users has been negative and characteristic of a ‘complaint tradition’ (Bolton, 2003). There is a long-standing belief that any distinctive China English expression must by its nature be bad English.

On the attitudes of people in China towards English, He and Li (2009) found that the majority of the respondents showed a preference for standard, inner circle varieties, especially in formal communication. Such attitudes may change, however, if CE features are gradually incorporated into the national ELT curriculum. There is good prospect that when a shift in attitudes occurs, societal demand for clearer delineation of CE features may increase, hence a healthy cycle how CE research may impact on CE pedagogy, and vice versa. As D. C. S. Li (2007, p. 14) proposes, the primary objectives of teaching English in China and Hong Kong should be to enable learners to communicate effectively, without compromising their L1 identity, and at the same time to encourage students to be proud and confident of their own localised English varieties.

Learning English and using it in communication worldwide should not oblige CE users to acquire inner circle English cultural practices along with the



language. The reason behind this argument would always be the need to maintain language identity, the fact that people need to signal to other people who they are and what lingua-cultural and/or social group(s) they wish or even take pride to be affiliated with when interacting with others. Research into CE helps establish language identity in international communication. China has never been a “language free community” (Makoni & Pennycook, 2007, p. 14).

The language ecology is such that CE users find themselves in a community where Chinese and English are default languages they need to know. How to maintain language identity when using English is thus a very important issue. Herein lies the significance of uncovering features of CE that are perceived to have an essential role to play towards achieving this goal. The stance of the ideal CE speaker/writer is one who is proud of his or her lingua-cultural identity, and who confidently enacts that identity through “Being Chinese, Speaking English” (Lo Bianco, Orton & Gao, 2009, p. 294).

The most important and revolutionary trend for English language teaching in the 21st century has been the choice of pedagogic model for the language classroom in countries like China. Considering the millions of students who learn English everyday as a compulsory course in school from primary (age 6-7) to university (age 22-23), the teaching model is extremely vital and meaningful.



The next step towards gaining wider acceptance and acknowledgment of CE as an autonomous variety is to adopt a more CE-based model for English language teaching. Therefore, it is proposed that CE research goals beset, and a CE-informed model ultimately be adopted, for the teaching and learning of English in China should appropriately reflect and be relevant to the norms and needs of local learners and users of English. Far from being neglected, CE features should have a place in the national ELT curriculum.

7.4 From lexical variants to collocational variants

The principal contribution of this thesis is that it complements previous studies on CE by taking a closer look at the larger picture. Many Chinese linguists have long noticed the lexical variants in CE. As a native Chinese Mainlander, I can easily think of examples like those in Table 7.2.



Table 7.2: Examples of CE lexis

Domains	Chinese	Pinyin	English
architecture	吊腳樓 祠堂	<i>diàojiǎolóu</i> <i>cítáng</i>	wooden house projecting over the water, ancestral temple
clothing	旗袍 唐裝	<i>qípáo</i> <i>tángzhuāng</i>	Cheongsam, a female dress with distinctive Chinese features and enjoys a growing popularity in the international world of high fashion
festival	春節 中秋	<i>chūnjié</i> <i>zhōngqiū</i>	Spring Festival, Mid-Autumn Festival
food	豆腐 雲吞	<i>dòufǔ</i> <i>húntún</i>	bean curd (tofu), wanton
cultural artefacts	喜糖 紅包	<i>xǐtáng</i> <i>hóngbāo</i>	Wedding Candy, Red Pocket (Lucky Money)
art	京劇 粵劇	<i>jīngjù</i> <i>yuèjù</i>	Peking Opera, Cantonese Opera
craft	繡花 皮影	<i>xiùhuā</i> <i>píyǐng</i>	Embroidery, Shadow puppet

Lexical expressions are needed when local, culture-specific items such as these are invoked or expressed in English. Take the word *hutong* as an example, people used to translate it as *street* or *lane*, and the like. However, more and more people realise that such *ad hoc* translations are imprecise and so later on, *hutong* was gradually accepted as an English word, a lexical borrowing from Mandarin Chinese, among others. In this case, flexibility in terms of legitimization of lexical variants is achieved.

Another example is that, in a popular film produced by American company DreamWorks Studios in 2008, *Kung Fu Panda*, the term ‘master’ and its equivalent in the original ‘*shifu*’ (Pinyin: *shīfū*) are both used. Their meaning is identical. They are both addressing someone who teaches you *kungfu* in the Chinese context. But the term ‘master’ is English and *shifu* is Mandarin Chinese. The acceptance and actual use of these two words alternatively, in a film targeting audiences worldwide, suggests that lexical variants and alternations of a Chinese-culture-specific word (addressing someone who teaches you *kung fu*) are achieved.

The point is that collocational variants of CE are also needed. For example, in her book 《中式英語之鑒》 (*Zhongshi yingyu zhi jian*, ‘Lessons Drawn from Others’ Mistakes in Chingish’), Pinkham (2000) listed many mistakes she had noted from various written sources where English was used in a Chinese context during the long period of time when she lived in China. For instance, the collocations *new innovations* and *family relative* are considered to be mistakes or improper English (Pinkham, 2000, p. 26). However, given that 22 tokens and 5 tokens of these two collocations respectively are found in BNC, as shown in the KWIC of *new INNOVATION* (Figures 7.1–7.2) and *family RELATIVE* (Figures 7.3–7.4), there seems no reason to label them as errors. On the contrary, they should be recognised as legitimate collocations.



r turning. Spring and foam mattresses are a fairly	new innovation	, arguably combining the best of both worlds. Most are rather
task of transplanting seedlings easier by a simple	new innovation	from Tuff-Link called the ‘Nursery Minder’. It is a
on on the subject. The Tinnitus Masker is a fairly	new innovation	which provides relief to many (but not all) sufferers.
things as us now, I have come up with a startling	new innovation	. The thermal beard. The beards come in a range of
inburgh-Brighton) and there is also — an entirely	new innovation	— an overnight sleeping-car service between Poole and Glasgow
looks very much like a Saab! They are promising	new innovations	on the safety and environmental fronts, as well as a wider
den — so get snapping! Mango to teddy bears —	new innovations	— from [gap:name] [gap:name] has launched two new muesli pr
he emphasis, Serratoso's book does contain some	new innovations	for teaching the difficult subject of the design of synthesis, and
e system, which assesses the cost effectiveness of	new innovations	before they are introduced and helps to rationalise their introduct
ventions and their implementation in the form of	new innovations	. Although Mensch's theory is in the same tradition as Schumpete
neurs see the opportunity to profit by investing in	new innovations	. From their example, people with less entrepreneurial ability will
e systems are working satisfactorily. Many of the	new innovations	have only been possible because of the company's commitment t
s being the most important methods of protecting	new innovations	, while secrecy and marketing-related advantages were generally
a such as earnings figures, future corporate plans,	new innovations	, or inventions, and would, as a result, owe
i has had to become more adaptable to successive	new innovations	. Looking back over this period most farm workers view the intro
es and polytechnics. We will continue to develop	new innovation	schemes for small and medium-sized businesses, including the h
s invention. But, despite this ability to pick up on	new innovations	, and add refinements of their own, neither Paul nor any
22(15cm) and in six different colours, it was a	new innovation	. In that year they introduced their ‘New Dwarf Blotched and
with maximum power. DPW is combined with a	new innovation	— the Symmetrical Bridge Concept (SBC). The throat bridge
ot very popular, they will probably get better, and	new innovations	will follow thick and fast. Since tomorrow's materials will be
, subscription are also included on your invoice, a	new innovation	, so that ALL fees may be paid together. It is
safety regulations operated by the Suez Canal. His	new innovation	won approval and orders for the first 10 tankers were placed with

Figure 7.1: KWIC of *new INNOVATION* in BNC

is the westerners who become the butt of the jokes. The	<u>family relatives</u>	in Bavaria — fat, bigoted, suburban and narrow-minded — are
common consent. Another 25 per cent of abusers were close	<u>family relatives</u>	such as brothers, uncles and grandfathers. The type and duration
re are some possible answers. One possible answer is that the	<u>family relatives</u>	of the eaten butterfly may benefit from its death. Members of
£645 a year. The value of child benefit for a three-child	<u>family relative</u>	to average earnings has fallen by 33 per cent. over a
ould mean something altogether different because there close	<u>family relatives</u>	take absolute priority and, no matter how important other business

Figure 7.3: KWIC of *family RELATIVE* in BNC

ly relatives frustrate? We are family relatives, just off the plane to Melbourne, and his face
 ager lady Chia, Mrs. Hsueeh, and the family relative, Mrs. Li, present Wen Pao this money to purchase
 places placed around their necks by a family relative. Be fond of quarreling with the relative in
 it neither, my only think of and each family relative stops every day joyfully under the same eave to
 the following order: Talking with a family relative about it (17.3%), seek support on the internet

Figure 7.4: KWIC of *family RELATIVE* in CEC

As shown in Figure 7.1 and 7.2, the collocation *new INNOVATION* in BNC are used to describe facilities, models or inventions that are recently designed. To the extent that new inventions crop up every now and then around the world, as an expression in reference to recent innovations, *new INNOVATION* is clearly well motivated. In CEC, however, the meaning of *new INNOVATION* is extended and stretched to include ‘recently developed ideas and motivations’. So the same collocation, *new INNOVATION*, may or may not mean the same thing in BrE and CE. On the other hand, with regard to the collocation *family RELATIVE*, as shown in Figures 7.3–7.4, there does not seem to be any noticeable difference in meaning across the BrE and CE tokens, suggesting that the collocation is considered quite acceptable by BrE users.

Pinkham’s (2000) view, that *new innovation* and *family relatives* are errors in need of correction, is therefore open to doubt. Now if longer collocations are scrutinised for a similar comparison and contrast, I feel certain that more subtle usage differences across BrE and CE may be identified. In my view, this is a promising area of further research into the distinctive features of CE. Through this thesis, I hope to have laid the groundwork for this future direction of research.

7.5 Limitations

Owing to the constraints of time and space, there are two somewhat obvious limitations from the point of view of methodological rigour. First, the corpora, CEC and BNC, only allow us to observe what is represented in them; what is not represented will be obscure to the investigator, especially when no spoken data was available for confirmation and comparison. Second, it would have been more revealing if longer collocations, up to four words, in both BrE and CE had been investigated more systematically.

7.6 Future research

The prime motivation for researching CE collocational features was that, seen in the light of what we know about CE lexis, listing individual words is inadequate. By identifying collocations, both overt and covert, there is good potential for better understanding how CE users' lexico-grammatical preferences help them enact their identity as CE users. However, this is only the first step towards identifying a fuller picture of collocational features of CE. As stated in Section 3.2, collocations (in the sense of the frequency of co-occurrence of words), should by no means be limited to two-word combinations. Rather, longer words or patterns should also be investigated in



future research. It is hoped that the methodological design of this study will be followed by larger-scale research on the collocational patterns of educated CE users in terms of a wider range of CE collocational patterns.

7.7 Coda

Four research questions were listed in Chapter 2 and addressed in subsequent chapters of this thesis. It is hoped that the methodological approach has been fruitful, the data analysis together with the ensuing argumentation thorough, and that the findings have enhanced our understanding of the many complexities surrounding research on China English.

By way of ending this thesis, I would like to share some interesting and highly relevant images by a designer, whose purpose is to create enlightening and stimulating contrasts regarding systematic, if stereotypical differences between broadly conceived *Chinese ways* and *Western ways* of doing and seeing things (Figure 7.5).



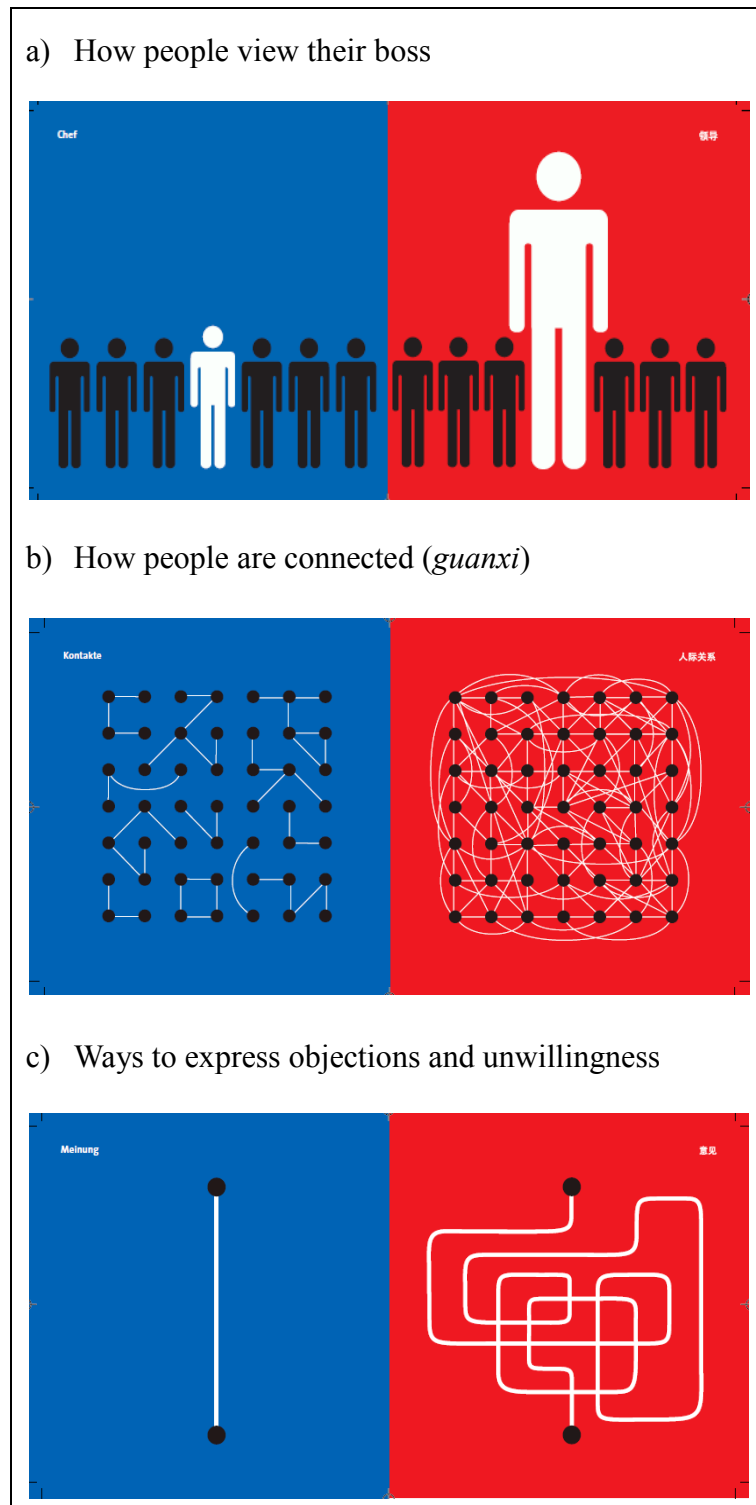


Figure 7.5: Doing and seeing things the Chinese (red) and western (blue) way

(Y. Liu, 2013)

These images encapsulate diametrically opposed, partly culturally conditioned stances or responses that are so common among Chinese and ‘westerners’ in the three above-mentioned situations. People in China see themselves as humble beings in big groups; jí tǐ (集體, similar to group or community one belongs to) is always more important than individuals. Consequently, Mainland Chinese people view their boss or leader (lǐngdǎo, 領導) in this group as the person who has the right to control them rather than just lead them in their work. A personal network or *guanxi* gradually and naturally emerges, in a manner that is more often than not complicated and that pervades people’s everyday lives (Y. Gu, 2002). Being humble and controlled, Mainland Chinese people are not so passionate in verbalising feelings, especially negative feelings towards each other. It is not that they don’t have feelings; rather, they tend not to show them in indirect ways. Indirectness as a distinctive trait for Chinese culture is manifested and perceptible in their ways of expressing objections, which have been passed on from generation to generation for thousands of years. All this is lingua-culturally reflected in the Chinese language. The CE collocational features as evidenced in this thesis are thus one distinct and important ‘different’ (rather than ‘deficit’) feature that CE users have every right to maintain in their communication with others, both intra- and internationally.

With the rapid development today of internet communication activities using electronic products that almost everyone in the world has at hand, platforms for information exchange are no longer reserved for the rich and powerful. With mobile phones and other electronic gadgets becoming more and more popular and accessible, people involved in such communication need to recognise clearly that the cultural 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 recognised. I sincerely hope that this study can contribute to a better understanding of the nature of the challenge towards identifying the collective lexico-grammatical preferences of Mainland Chinese when they are using English with others, at home or across borders.



References

- Abercrombie, D. (1965). *Studies in phonetics and linguistics*. Oxford: Oxford University Press.
- Adamson, B. (2002). Barbarian as a foreign language: English in Chinese schools. *World Englishes*, 21(2), 231-43.
- Adamson, B. (2004). *China's English: a history of English in Chinese education*. Hong Kong: Hong Kong University Press.
- Adamson, B., Bolton, K., Lam, A. & Tong, Q. S. (2002). English in China: a preliminary bibliography. *World Englishes*, 21(2), 349-355.
- Adamson, B., & Feng, A. (2014). Models for trilingual education in the People's Republic of China. In D. Gorter, V. Zenotz, & J. Cenoz, (Eds.), *Minority languages and multilingual education: bridging the local and the global* (pp. 29-44). Dordrecht: Springer.
- Aijmer, K. (2009). *Corpora and language teaching*. Amsterdam: John Benjamins.
- Aijmer, K. & Altenberg, B. (1991). *English corpus linguistics: studies in honour of Jan Svartvik*. London: Longman.
- Aijmer, K. & Altenberg, B. (2013). *Advances in corpus-based contrastive linguistics*. Amsterdam: John Benjamins.
- Anthony, L. (2011). AntConc (Version 3.2.1) [Computer Software]. Tokyo: Waseda University.
- Aston, G., & Burnard, L. (1998). *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Bamgbose, A. (1998). Torn between the norms: innovations in World Englishes. *World Englishes*, 17(1), 1-14.
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Belcher, D. D., Johns, A. M., & Paltridge, B. (2011). *New directions in English for specific purposes research*. Ann Arbor: University of Michigan Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Block, D. (2007). The rise of identity in SLA research, post Firth and Wagner (1997). *The Modern Language Journal*, 91(5), 863-876.
- Blommaert, J. (2010). *The sociolinguistics of globalisation*. Cambridge: Cambridge University Press.
- Bloome, D., & Green, J. (2002). Directions in the sociolinguistic study of reading. In P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 395-421). Mahwah, NJ: Lawrence Erlbaum.
- Bolton, K. (2000). The sociolinguistics of Hong Kong and the space for Hong



- Kong English. *World Englishes*, 19(3), 265-85.
- Bolton, K. (2002). *Hong Kong English: autonomy and creativity*. Hong Kong: Hong Kong University Press.
- Bolton, K. (2002). Chinese Englishes: from Canton jargon to global English. *World Englishes*, 21(2), 181-99.
- Bolton, K. (2003). *Chinese Englishes: a sociolinguistic history*. Cambridge: Cambridge University Press.
- Bolton, K. (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today*, 24(2), 3-12.
- Bolton, K. (2012). World Englishes and linguistic landscapes. *World Englishes*, 31(1), 30-33.
- Bolton, K. & Graddol, D. (2012). English in China today. *English Today*, 28(1), 3-9.
- Bolton, K., Graddol, D., & Meierkord, C. (2011). Towards developmental world Englishes. *World Englishes*, 30(4), 459-480.
- Bolton, K., & Kachru, B. B. (2006). *World Englishes*. New York: Routledge.
- Bolton, K., & Tong, Q. S. (2002). Special issue on English in China: Interdisciplinary perspectives. *World Englishes*, 21(2), 177-355.
- Bond, M. H. (1996). *The handbook of Chinese psychology*. New York: Oxford University Press.
- Botha, W. (2013). English-medium instruction at a university in Macau: Policy and realities. *World Englishes*, 32(4), 461-475.
- Brown, G., Malmkjaer, K. & Williams, J. (1996). *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press.
- Brown, K. & Miller, J. (2013). *The Cambridge Dictionary of Linguistics*. Cambridge: Cambridge University Press.
- Butler, S. (1997). Corpus of English in Southeast Asia: implications for a regional dictionary. In M. L. Bautista (Ed.), *English is an Asian language: the Philippine context* (pp. 103-124). Sydney: Macquarie Library.
- Cai, J. G. (2004). Daxue yingyu jiaoxue fazhan shishang de liangge xinde tupo [Two breakthroughs in history in reforms in college English teaching]. *Zhongguo waiyu [Foreign Languages in China]*, 01, 19-22.
- Cameron, D. (1995). *Verbal hygiene*. London: Routledge.
- Canagarajah, S. (2004). *Reclaiming the local in language policy and practice*. Mahwah: Lawrence Erlbaum.
- Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal*, 91(5), 923-939.
- Cannon, G. (1988). Chinese borrowings in English. *American Speech*, 63, 3-33.
- Carter, R. (2012). *Vocabulary: applied linguistic perspectives*. New York: Routledge.



- Čermák, F. (2003). Today's corpus linguistics: some open questions. *International Journal of Corpus Linguistics*, 7(2), 265-282.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen, X. (2010). *Discourse-grammatical features in L2 speech: a corpus-based contrastive study of Chinese advanced learners and native speakers of English*. PhD dissertation, City university of Hong Kong.
- Cheng, C. C. (1992). Chinese varieties of English. In B. B. Kachru (Ed.), *The other tongue: English across Cultures* (pp. 162-77). Urbana, IL: University of Illinois Press.
- Clyne, M. (1992). *Pluricentric languages: differing norms in different nations*. Berlin: Mouton de Gruyter.
- Clyne, M. & Sharifian F. (2008). English as an international language. *Australian Review of Applied Linguistics*, 31(3), 281-296.
- Cogo, A., & Martin D. (2012). *Analysing English as a lingua franca: a corpus-driven investigation*. London: Continuum.
- Conrad, A. W. (1996). The international role of English: the state of the discussion. In J. A. Fishman, A. W. Conrad & A. Rubal-Lopez (Eds.), *Post-imperial English: Status change in former British and American colonies 1840-1990* (pp.13-36), Berlin: Mouton de Gruyter.
- Cook, V. (1996). *Second language learning and language teaching*. London: Arnold.
- Cook, V. (2002). *Portraits of the L2 user*. Clevedon: Multilingual Matters.
- Corder, S. P. (1981). *Error analysis and interlanguage*. London: Oxford University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.
- Crowley, T. (2003). *Standard English and the politics of language* (2nd edn.). Houndmills, Basingstoke: Palgrave Macmillan.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Crystal, D. (2001). *Language and the internet*. Cambridge: Cambridge University Press.
- Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.
- Crystal, D. (2012). A global language. In P. Seargeant, & J. Swann, (Eds.), *English in the world: history, diversity, change* (pp. 151-177). London: Routledge.
- Cummings, P. J. & Wolf, H. G. (2011). *A dictionary of Hong Kong English. Words from the Fragrant Harbor*. Hong Kong: Hong Kong University Press.
- Cui, X. X. (2006). An understanding of 'China English' and the learning and



- use of the English language in China. *English Today*, 88(1), 40-43.
- Dai, F. (2010). English-language creative writing in mainland China. *World Englishes*, 29(4), 546-556.
- Dai, W. D. (2002). Dui waiyu jiaoxue 'yitiaolong' gaige de sikao [On the reform pattern of foreign language teaching], *Waiyu Jie [Foreign Language World]*, 11(01), 26-31.
- Dai, W. D. (2009). Zhongguo gaoxiao waiyu jiaoyu 30 nian [Foreign languages education in universities in China: a 30 years' review], *Waiyujie [Foreign Language World]*, 130(01), 2-4.
- Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English: word sketches, collocates and thematic lists*. New York: Routledge.
- Davydova, J. (2012). English in the outer and expanding circles: a comparative study. *World Englishes*, 31(3), 366-385.
- de Swaan, A. (2001). *Words of the world: the global language system*. Cambridge: Polity Press.
- de Swaan, A. (2010). Language systems. In: N. Coupland (Ed.), *The handbook of language and globalisation* (pp. 57-76). Malden: Wiley-Blackwell.
- Delbridge, A. (1981). *The Macquarie Dictionary*. Sydney: Macquarie Dictionary Publishers.
- Delbridge, A. (1999). Standard Australian English. *World Englishes*, 18(2), 259-270.
- Deterding, D. (2006). The pronunciation of English by speakers from China. *English World-Wide*, 27, 175-98.
- Ding, G. X. (1985). *A new Chinese-English dictionary*. Hong Kong: Joint Publishing.
- Dixon, R. M. (2011). *The languages of Australia*. Cambridge: Cambridge University Press.
- Du, R., & Jiang, Y. (2001). Jin'er shi nian 'zhongguoyingyu' yanjiu shuping ['China English' in the past 20 years]. *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, 33(1), 37-41.
- Du, Z. M. (1998a). Shijie yingyu yu jingyu zhongguo yingyu zhong de yuyan yu wenhua [World Englishes and language and culture in China English]. *Waiyu yu waiyujiaoxue [Foreign Languages and Foreign Languages Teaching]*, 110(08), 14-16.
- Du, Z. M. (1998b). Zhongguo yingyu wenti ji qita [Issues relate to China English research]. *Waiyu jiaoxue [Foreign Language Teaching]*, 77(3), 6-14.
- Dudley Evans, T. & St. John, M. J. (1998). *Developments in English for specific purposes: A multidisciplinary approach*. Cambridge: Cambridge University Press.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford

University Press.

- Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, 4(3), 193-220.
- Evans, S. (2011). Hong Kong English and the professional world. *World Englishes*, 30(3), 293-316.
- Facchinetti, R. (2007). *Corpus linguistics 25 years on*. Amsterdam: Rodopi.
- Feng, Z. W. (2007). *The role of English language in China*. Retrieved from <http://www.docin.com/p-456068752.html>. Accessed 10 July 2014.
- Fenglin (2005). Text-organizer [software]. Retrieved from <http://www.skycn.com/soft/appid/5633.html>. Accessed 10 July 2014.
- Firth, A. & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal*, 81(2), 286-300.
- Firth, A., & Wagner, J. (2007). Second/foreign language learning as a social accomplishment: elaborations on a reconceptualised SLA. *The Modern Language Journal*, 91(s1), 800-819.
- Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics 1934-1951* (pp. 190-215). Oxford: Oxford University Press.
- 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.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Friginal, E., & Hardy, J. (2014). *Corpus-based sociolinguistics: a guide for students*. New York: Routledge.
- 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.
- Gao, X. S. (2012). The Study of English in China as a patriotic enterprise. *World Englishes*, 31(3), 351-365.
- Gao, Y. H. (2004). *Cultural nationalism in contemporary China: the search for national identity under reform*. New York: Routledge.
- Gao, Y. H., Cheng Y., Zhao Y., & Zhou Y. (2005). Self-identity changes and English learning among Chinese undergraduates. *World Englishes*, 24(1), 39-51.
- Gass, S. M. (1998). Apples and Oranges: or, why apples are not orange and don't need to be: a response to Firth and Wagner. *The Modern Language Journal*, 82(1), 83-90.
- Gass, S. M., Lee, J. & Roots, R. (2007). Firth and Wagner (1997): new ideas or anew articulation? *The Modern Language Journal*, 91(S1), 788-799.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.



- Ge, C. G. (1980). Mantan you hanyiyingwen [Random thoughts on some problems in Chinese-English translation], *FanyiTongxun [Chinese Translator's Journal]*, 2, 1-8.
- Gee, J. P. (1999). *An introduction to discourse analysis: theory and method*. London: Routledge.
- Giles, H. & Powesland, P. (1975). *Speech style and social evaluation*. London: Academic Press.
- Gilquin, G. & Gries, S. (2009). Corpora and experimental methods: a state-of-the-artreview. *Corpus Linguistics and Linguistic Theory*, 5(1), 1-26.
- Gimson, A. C. (1980). *An introduction to the pronunciation of English*. London: Edward Arnold.
- Görlach, M. (1990). *Studies in the history of the English language*. Heidelberg: Carl Winter.
- Görlach, M. (1991). *Varieties of English around the world*. Amsterdam: John Benjamins.
- Görlach, M. (2002). *Still more Englishes*. Amsterdam: John Benjamins.
- Graddol, D. (1997). *The future of English*. London: British Council.
- Graddol, D. (2006). *English next*. London: British Council.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation. In K. Aijmer, (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam: John Benjamins.
- Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units? In A.O'Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 212-226). New York: Routledge.
- Greenbaum, S. (1996). *Comparing English worldwide: the international corpus of English*. Oxford: Clarendon.
- Gregg, K. R. (2003). SLA theory: Construction and assessment. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 831-865). Malden, MA: Blackwell.
- Gries, S. T. (2013). 50-something years of work on collocations: what is or should be next.... *International Journal of Corpus Linguistics*, 18(1), 137-166.
- Gries, S. T., & Joybrato, M. (2010). Lexical gravity across varieties of English: an ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics*, 15(4), 520-548.
- Gu, Y. G. (2002). Guanxi: backdoor practice as goal-directed discourse. In D. C. S. Li (Ed.), *Discourses in search of members* (pp. 273-298). Lanham: University Press of America.
- Gui, S. (1988). *Yingyong yingyu yu zhongguo yingyu jiaoxue [Applied English and English language teaching in China]*. Jinan, China: Shandong

- jiaoyuchubanshe [Shandong Education Press].
- Günthner, S. (2002). Moral work in intercultural communication: explicit reference to moral norms in German-Chinese interaction. In D. C. S. Li (Ed.), *Discourses in search of members* (pp. 361-391). Lanham: University Press of America.
- Haberland, H. (2013). ELF and the bigger picture. *Journal of English as a Lingua Franca*, 2(1), 195-198.
- Haberland, H., & Mortensen, J. (2012). Language variety, language hierarchy, and language choice in the international university. *International Journal of the Sociology of Language*, 216, 1-6.
- 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: Longmans, Green & Co. Ltd.
- Han, L. (2007). Zhongguo yingyu yanjiu xianzhuang fenxi [Current research on China English]. *Waiyu yu waiyu jiaoxue [Foreign Languages and Their Teaching]*, 223(10), 28-32.
- He, D. Y., & Li, C.S. (2009). Language attitudes and linguistic features in the 'China English' debate. *World Englishes*, 28(1), 70-89.
- He, D.Y., & Miller, L. (2011). English teacher preference: the case of China's non-English-major students. *World Englishes*, 30(3), 428-443.
- Heath, S. B. (1977). Language and politics in the United States. In M. Saville-Troike (Ed.), *Georgetown University round table on language and linguistics*. Washington: Georgetown University Press.
- Hickey, R. (2012). Standard English and standards of English. In R. Hickey (Ed.), *Standard of English: Codified varieties around the world* (pp. 1-33). Cambridge: Cambridge University Press.
- Ho, Y. (1993). Aspects of discourse structure in Mandarin Chinese. New York: Mellen University Press.
- Holmes, J. (2013). *An introduction to sociolinguistics*. London: Routledge.
- Hu, S. Z. (1993). *Yinghan xiuci bijiao yanjiu [Comparative studies in English and Chinese rhetoric]*. Shanghai: shanghai waiyu jiaoyu chubanshe [Shanghai foreign language education press].
- Hu, W. (1995). *Functional perspectives and Chinese word order*. PhD dissertation, Ohio State University.
- Hu, X. L. (2008a). Zhongguo yingyu zhi gongneng yanjiu [On the function of China English]. *Waiyu yu waiyu jiaoxue [Foreign Languages and Their Teaching]*, 236(11), 13-16.
- Hu, X. L. (2008b). Zhongguo yingyu bianti yanjiu xianzhuang ji qianjing [Studies of China English: review and prospects]. *Zhongguo waiyu [Foreign Languages in China]*, 22(2), 27-30.
- Hu, X. Q. (2004). Why China English should stand alongside British, American, and the other 'world Englishes'. *English Today*, 20(2), 26-33.

- Hu, X. Q. (2005). China English, at home and in the world. *English Today*, 21(3), 27-38.
- Huang, J. Q. (1988). Yingdang kending 'xiyihanhua' xianxiang de jijimian [The positive role of Sinicism in the English-translated version]. *Zhongguo Fanyi [Chinese Translator's Journal]*, 1, 39-47.
- Huang, Y. H., & Miao, L. (2013). *Annual report on the development of China's study abroad 2013*. Beijing: Social Sciences Academic Press.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Husserl, E. (1970). *The crisis of European sciences and transcendental philosophy* (trans. by D.Carr). Evanston: Northwestern University Press.
- Hutchinson, T. & Waters, A. (1987). *English for specific purposes: a learning-centred approach*. Cambridge: Cambridge University Press.
- Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for academic purposes*, 1(1), 1-12.
- Jenkins, J. (2000). *The phonology of English as an international language: new models, new norms, new goals*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23(1), 83-103.
- Jenkins, J. (2003). *World Englishes: a resource book for students*. London: Routledge.
- Jenkins, J. (2006). Points of view and blind spots: ELF and SLA. *International Journal of Applied Linguistics*, 16(2), 137-162.
- Jenkins, J. (2007). *English as a lingua franca: attitude and identity*. Oxford: Oxford University Press.
- Jia, D. L. (1990). Siwei moshi yu xianxing xulie: hanshi yingyu yuxu tese [Thinking mindset and linear sequence: word-order in Sinicized English]. *Waiguo Yu [Foreign Languages]*, 69(5), 12-17.
- Jia, G. J., & Xiang, M. Y. (1997). Wei zhongguo yingyuyibian [In defense of Chinese English]. *Waiyu Yu Waiyu Jiaoxue [Foreign Languages and Foreign Language Teaching]*, 105(5), 11-12.
- Jiang, W. Y. (2009). *Acquisition of word order in Chinese as a foreign language*. Berlin: Mouton de Gruyter.
- Jiang, Y. J., & Du, R. Q. (2003). Youguan 'zhongguoyingyu' de wenti [Issues on 'China English']. *Waiyu Jiaoxue [Foreign Language Education]*, 24, 27-35.
- Jin, H. K. (2002). Hanying kuawenhua jiaoji fanyi zhong de zhongguo yingyu [China English in the intercultural translation between Chinese and English]. *Guangdong Zhiye Jishu Shifa Xueyuan Xuebao [Journal of Guangdong Polytechnic Normal University]*, 2, 72-78.

- Jones, S., & Sinclair, J. M. (1974). English lexical collocations: a study in computational linguistics. *Cahiers de lexicologie*, 24, 15-61.
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kachru, B. B. (1982). *The Other Tongue: English across Cultures*. Oxford: Pergamon. Second edition (1992), Urbana, IL: University of Illinois Press.
- Kachru, B. B. (1983). *The Indianisation of English*, New Delhi: Oxford University Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.
- Kachru, B. B. (1986). *The alchemy of English: the spread, functions, and models of non-native Englishes*. Oxford: Pergamon Press.
- Kachru, B. B. (1990). World Englishes and applied linguistics. *Studies in Linguistic Sciences*, 19(1), 127-52.
- Kachru, B. B. (1996). World Englishes: agony and ecstasy. *Journal of Aesthetic Education*, 30(2), 135-55.
- Kachru, B. B. (2005). *Asian Englishes: beyond the canon*. Hong Kong: Hong Kong University Press.
- Kachru, Y. & Nelson C. L. (2006). *World Englishes in Asian contexts*. Hong Kong: Hong Kong University Press.
- Kasper, G. (1997), 'A' stands for acquisition: a response to Firth and Wagner. *The Modern Language Journal*, 81(3), 307-312.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Kennedy, G. (2003). *Structure and meaning in English: a guide for teachers*. Harlow: Longman.
- Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 21-41). Amsterdam: John Benjamins.
- Kenyon, J. S. (1924). *American pronunciation*. Ann Arbor: George Wahr.
- Kilgarriff, A. (2006). Collocationality (and how to measure it). In *Attidel XII Congresso Internazionale di Lessicografia: Torino, 6-9 September 2006* (pp. 997-1004).
- Kirkpatrick, A. (2006). Which model of English: native speaker, nativised or lingua franca? In R. Rubdy, & M. Saraceni (Eds.), *English in the world: global rules, global roles* (pp. 71-83), London: Continuum Press.
- 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*. London: Routledge.
- Kirkpatrick, A., & Xu, Z. C. (2002). Chinese pragmatic norms and 'China English'. *World Englishes*, 21(2), 269-280.
- Kloss, H. (1978). *Die Entwicklung neuer germanischer Kultursprachen seit 1800* (2nd edn.). Düsseldorf: Schwann.
- Kramsch, C. & Whiteside, A. (2007). Three fundamental concepts in second language acquisition and their relevance in multilingual contexts. *The Modern Language Journal*, 91(5), 907-922.
- Lam, A. (2002). English in education in China: policy changes and learners' experiences. *World Englishes*, 21(2), 245-256.
- Lantolf, J. P. (2007). Extending Firth and Wagner's (1997) ontological perspective to L2 classroom praxis and teacher education. *The Modern Language Journal*, 91(5), 877-892.
- Larsen-freeman, D. (2007). Reflecting on the cognitive-social debate in second language acquisition. *The Modern Language Journal*, 91(S1), 773-787.
- Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Leech, G. N. (1966). *English in advertising: a linguistic study of advertising in Great Britain*. London: Longman.
- Leech, G. N. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in corpus linguistics: proceedings of Nobel Symposium 82* (pp. 105-122), Stockholm, 4-8 August 1991. Berlin: Mouton de Gruyter.
- Leech, G. N. (2002). The importance of reference corpora. Keynote speech at *Hizkuntza-corpusak. Oraina eta geroa*. (2002-10-24/25).
- Leech, G. N. (2011). Why frequency can no longer be ignored in ELT. *Foreign language Teaching and Research*, 43(1), 3-20.
- Leech, G. N., Deuchar, M., & Hoogenraad, R. (2006). *English grammar for today: a new introduction* (2nd edn.). Basingstoke England: Palgrave Macmillan.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.
- Leech, G., & Svartvik, J. (2013). *A communicative grammar of English*. New York: Routledge.
- Leung, H. M. & Li, D. C. S. (2011). The state of Multilingualism in Macau SAR: An Update. *Chinese Cross Currents*, 8(1), 82-96.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference*

- grammar*. Berkeley: University of California Press.
- Li, David C. S. (1999). The functions and status of English in Hong Kong: a post-1997 update. *English World-Wide*, 20(1), 67-110.
- Li, David C. S. (2000). Hong Kong English: new variety of English or interlanguage. *English Australia Journal*, 18(1), 50-59.
- Li, David C. S. (2001). L2 lexis in L1: reluctance to translate out of concern for referential meaning. *Multilingual*, 20(1), 1-26.
- Li, David C. S. (2002). Pragmatic dissonance: the ecstasy and agony of speaking like a native speaker of English. In D. C. S. Li (Ed.), *Discourses in search of members* (pp. 559-594). Lanham: University Press of America.
- Li, David C. S. (2006). Problematising empowerment: on the merits and demerits of non-native models of English in the EIL curriculum. *Southeast Asia: A Multidisciplinary Journal*, 6(1), 112-31.
- Li, David C. S. (2007). Researching and teaching China and Hong Kong English. *English Today*, 23(3-4), 11-17.
- Li, David C. S. (2009). Researching non-native speakers' views toward intelligibility and identity: bridging the gap between moral high grounds and down-to-earth concerns. In F. Sharifian (Ed.), *English as an International Language: Perspectives and Pedagogical Issues* (pp.81-118). Bristol: Multilingual Matters.
- Li, David C. S. (2010a). When does an unconventional form become an innovation? In A. Kirkpatrick (Ed.), *The Routledge handbook of world Englishes* (pp. 617-33). London: Routledge.
- Li, David C. S. (2010b). Improving the standards and promoting the use of English in Hong Kong: issues, problems and prospects. In Anwei Feng (Ed.), *English language education across Greater China* (pp. 95-113). Clevedon: Multilingual Matters.
- Li, J. Y. (2009). Motivational force and imagined community in 'Crazy English'. In J. Lo Bianco, J. Orton & Y. H. Gao (eds.), *China and English. Globalisation and the dilemmas of identity* (pp. 211-223). Bristol, UK: Multilingual Matters.
- Li, L. Y., Zhang, R. S., & Liu, L. (1988). *Zhongguo yingyu jiaoxueshi [History of English teaching in China]*. Shanghai: Shanghai Foreign Language Education Press.
- Li, W. Z. (1993). *Zhongguo yingyu yu zhongguoshi yingyu [China English and Chinglish]. Waiyu Jiaoxue Yu Yanjiu [Foreign Languages Teaching and Research]*, 96(4), 18-24.
- Li, W. Z. (2005). *Extracting the multiword units from China's English news articles*. Paper presented at the BAAL Corpus Linguistics SIG/OTA Workshop: Identifying and Researching Multiword Units, (21st April 2005). Computing services, Oxford University.



- Li, W. Z. (2011). China English Corpus. [Electronic data].
- Li, Y. X. (ca. 1661-1722). *Di ZiGui [Standards for being a good pupil and child]*.
- Lian, S. N. (2010). *Hanying duibi yanjiu [Contrastive studies of English and Chinese]*. Beijing: Gaodengjiaoyuchubanshe [Higher Education Press].
- Liang, M. C., Xu, J. J., & Li, W. Z. (2010). *Yuliaoku yingyong jiaocheng [Using corpora: a practical coursebook]*. Beijing: Waiyu jiaoxue yu yanjiu chubanshe [Foreign language teaching and research press].
- Liao, X. Y. (2013). *Jiucheng daxuesheng ruxue jiazhang husong [90 percent university freshmen were escorted to school by parents in their first day]*. Chongqing shangbao [Chongqing Economic Times]. 17th September.
- Lin, Q. Y. (1998). Zuowei waiyu de yingyu bianti: zhongguo yingyu [China English: a variety as a foreign language]. *Waiyu yu waiyu jiaoxue [Foreign Languages and Their Teaching]*, 110(8), 16-17.
- Liu, R. Q. (1995). *Xifang yuyan xue liupai [Schools of Linguistics]*. Beijing: Waiyu jiaoxue yu yanjiu chubanshe [Foreign language teaching and research press].
- Liu, Y. (2013, 9th Print). *'Ost trifft West' [East Meets West]*. Mainz, Germany: Hermann Schmidt Verlag.
- Liu, Y. H. (2001). *Shiyong xiandai hanyu yufa [Practical grammar for Chinese language]*. Beijing: Shangwuyinshuguan [Commercial Press].
- Lo Bianco, J., Orton, J. & Gao, Y. H. (2009). *China and English: Globalisation and the dilemmas of identity*. Bristol: Multilingual Matters.
- Long, M. H. (1997). Construct validity in SLA research: a response to Firth and Wagner. *The Modern Language Journal*, 81(3), 318-323.
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli, (Eds.), *Text and technology: in honour of John Sinclair* (pp. 157-176). Amsterdam: John Benjamins.
- Louw, B., & Chateau, C. (2010). *Semantic prosody for the 21st century: Are prosodies smoothed in academic contexts? A contextual prosodic theoretical perspective*. Proceeding of paper presented in 10th International Conference on Statistical Analysis of Textural Data (pp. 755-764), JADT.
- Lü, S. X. (1990). *Zhongguo wenfa yaolue [Key grammars for Chinese language]*. Beijing: Shangwu yinshuguan [Commercial Press].
- Luo, Y. Z. (1998). Zhongguo yingyu qianjing guan [Perspectives in China English]. *Waiyu yu waiyu jiaoxue [Foreign Languages and Their Teaching]*, 107(5), 24-25, 51.
- Mackey, W. F. (1967). *Bilingualism as a world problem*. Montreal: Harvest

- House.
- Makoni, S., & Pennycook, A. (2007). *Disinventing and reconstituting languages*. Clevedon: Multilingual Matters.
- Mathews, R. H. (1960). *Mathew's Chinese-English dictionary*. Cambridge: Harvard University Press.
- Matthews, P. H. (2006). *Oxford concise dictionary of linguistics*. Shanghai: Shanghai Foreign Language Education Press.
- McArthur, T. (1987). The English Languages? *English Today*, 3(3), 9-13.
- McArthur, T. (1998). *The English Languages*. Cambridge: Cambridge University Press.
- McArthur, T. (2002). *Oxford Guide to World English*. Oxford: Oxford University Press.
- McArthur, T. (2003). World English, Euro-English, Nordic English? *English Today*, 19(1), 54-58.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M., & O'Dell, F. (2005). *English collocations in use. How words work together for fluent and natural English: self-study and classroom use*. Cambridge: Cambridge University Press.
- McCarthy, M., & O'Dell, F. (2008). *Academic vocabulary in use: 50 units of academic vocabulary reference and practice, self-study and classroom use*. Cambridge: Cambridge University Press.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Abingdon: Routledge.
- McIntosh, A., & Halliday, M. A. K. (1967). *Patterns of language: papers in general, descriptive and applied linguistics* (2nd edn.). London: Longmans.
- McKay, S. L. (2003). Toward an appropriate EIL pedagogy: re-examining common ELT assumptions. *International Journal of Applied Linguistics*, 13(1), 1-22.
- McKay, S. L. (2005). Teaching the pragmatics of English as an international language. *Guidelines*, 27(1), 3-9.
- McKay, S. L. (2009). English as an international language: Where we are and where we need to go. *Anthology Series–Seameo Regional Language Centre*, (50), 146-170.
- Medgyes, P. (1994). *The non-native teacher*. London: Macmillan.
- Melchers, G., & Philip S. (2011). *World Englishes* (2nd edn.). London: Hodder Education.
- Mesthrie, R. & Bhatt, R. M. (2008). *World Englishes: The study of new*



- linguistic varieties*. Cambridge: Cambridge University Press.
- Mesthrie, R. & Rakesh, B. (2008). *World Englishes: an introduction to new language varieties*. Cambridge: Cambridge University Press.
- Meyer, C.F. (2002). *English corpus linguistics*. Cambridge: Cambridge University Press.
- Milroy, J. (1999). The consequences of standardisation in descriptive linguistics. In T. Bex, & R. Watts, (Eds.), *Standard English: the widening debate* (pp. 16-39). London: Routledge.
- Milroy, L. (1999). Standard English and language ideology in Britain and the United States. In T. Bex, & R. Watts, (Eds.), *Standard English: the widening debate* (pp. 173-206). London: Routledge.
- Milroy, J., & Milroy, L. (1985). *Authority in language: investigating language prescription and standardisation*. (2nd edn.), London: Routledge.
- Ministry of Education. (1986). *The College English curriculum*. Shanghai Foreign Language Education Press.
- Moody, A. (2008). Macau English: status, functions and forms. *English Today*, 24(3), 3-15.
- Mori, J. (2007). BorderCrossings? Exploring the intersection of second language acquisition, conversation analysis, and foreign language pedagogy. *The Modern Language Journal*, 91(5), 849-862.
- Morris, M. (2004). Words of the World: The Global Language System by Abram de Swaan. *Language in Society*, 33 (4), 620-624.
- Mufwene, S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mukherjee, J., & Gries, S. T. (2009). Collostructional nativisation in New Englishes: verb-construction associations in the International Corpus of English. *English World-Wide*, 30(1), 27-51.
- Murata, K., & Jenkins, J. (2009). *Global Englishes in Asian contexts: current and future debates*. Houndmills, England: Palgrave Macmillan.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 3-13). Amsterdam: John Benjamins.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, 63(1), 59-81.
- Nation, I. S. P., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- National Bureau of statistics of China (2007). *China Statistical Yearbook*.

- Beijing: China Statistical Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nelson, G., Wallis, S., & Aarts, B. (2002). *Exploring natural language: working with the British component of the international corpus of English*. Amsterdam: John Benjamins.
- Nelson, C. L. (2011). *Intelligibility in world Englishes: theory and application*. New York: Routledge.
- Niu, Q., & Wolff, M. (2003). China and Chinese, or Chingland and Chinglish. *English Today*, 19(4), 30-35.
- Norton, B. (1995). Social identity, investment, and language learning. *TESOL Quarterly*, 29(1), 9-31.
- Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly*, 31, 409-429.
- Norton, B. (2000). *Identity and language learning: gender, ethnicity and educational change*. Harlow: Longman.
- Nunan, D. & Carter, R. (2001). *The Cambridge guide to teaching English to speakers of other language*. Cambridge: Cambridge University Press.
- Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Pan, W.G. (2010). *Hanying yuyan duibi gailun [Introduction to ChineseEnglish contrastive studies]*. Beijing: Shangwu yinshuguan [Commercial Press].
- Pan, W.G., & Yang, Z. J. (2008). *Yinghan duibi de lilun yu fangfa yanjiu [Theory and practice on English-Chinese contrastive study]*. Shanghai: Shanghai waiyu jiaoyu chubanshe [Shanghai Foreign Language Education Press].
- Pang, J., Zhou, X. & Fu, Z. (2002). English for international trade: China enters the WTO. *World Englishes*, 21(2), 201-216.
- Partington, A. (1998). *Patterns and meanings: using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- 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.
- Pennycook, A. (2001). *Critical applied linguistics: a critical introduction*. Mahwah: Lawrence Erlbaum.
- Pennycook, A. (2007). *Global Englishes and transcultural flows*. London & New York: Routledge.
- Philip, G. (2011). *Colouring meaning: Collocation and connotation in figurative language*. Amsterdam: John Benjamins.
- Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.



- Phillipson, R. (2004). English in globalization: Three approaches. *Journal of Language, Identity & Education*, 3 (1), 73–78.
- Phillipson, R. (2009). Linguistic imperialism continued. London: Routledge.
- Pinkham, J. (2000). *Zhongshi yingyu zhi jian [Lessons Drawn from Others' Mistakes of Chingish]*. Beijing: Foreign Language Teaching and Research Press.
- Poullisse, N. (1997). Some words in defense of the psycholinguistic approach: a response to Firth and Wagner. *The Modern Language Journal*, 81(3), 324-328.
- Qiu, L. & Ning, Q. (2002). Zhongguo yingyu zhiyi [Challenge China English]. *Waiyu jiaoxue [Foreign Language Education]*, 23(6), 23-26.
- Quirk, R. (1995). *Grammatical and lexical variance in English*. London: Longman.
- Quirk, R. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- Ramson, W. S. (1988). *The Australian national dictionary: a dictionary of Australianisms on historical principles*. Melbourne: Oxford University Press.
- Ren, X.L. (1981). *Hanying bijiao yufa [A Chinese-English comparative grammar]*. Beijing: Zhongguo shehui kexue chubanshe [China social science 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.
- Renouf, A. & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 128-143). London: Longman.
- Robinson, P. (1991). *ESP today: A practitioner's guide*. Hemel Hempstead: Prentice Hall.
- Ross, H. A. (1993). *China learns English: language teaching and social change in the people's republic*. New Haven: Yale University Press.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Schmitz, J. R. (2014). Looking under Kachru's (1982, 1985) three circles model of World Englishes: the hidden reality and current challenges.



- Retrieved from http://www.scielo.br/scielo.php?pid=S1984-63982014005000010&script=sci_arttext. Accessed 11 July 2014.
- Schneider, E. W. (1997). Chaos theory as a model for dialect variability and change? In A. A. Thomas (Ed.), *Issues and methods in dialectology* (pp.22-36). Bangor: University of Wales, Department of Linguistics.
- Schneider, E. W. (2003). The dynamics of New Englishes: from identity construction to dialect birth. *Language*, 79(2), 233-281.
- Schneider, E. W. (2010). Developmental patterns of English: similar or different? In A. Kirkpatrick (Ed.), *The Routledge handbook of world Englishes* (pp. 372-84). London: Routledge.
- Schneider, E. W. (2011). *English around the world: an introduction*. Cambridge: Cambridge University Press.
- Schneider, E. W. (2013). English as a contact language: the ‘new Englishes’. In D. Schreier & M. Hundt (Eds.), *English as a contact language* (pp. 131-148). Cambridge: Cambridge University Press.
- Schneider, E. W. (2014). New reflections on the evolutionary dynamics of world Englishes. *World Englishes*, 33(1), 9-32.
- Scollon, R. & Scollon, S. B. K. (1995). *Intercultural communication: a discourse approach*. Malden: Blackwell.
- 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.
- Sergeant, P. (2012). *Exploring world Englishes: language in a global context*. London: Routledge.
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209-239.
- Seidlhofer, B. (2006). English as a lingua franca in the expanding circle: what it isn't. In R. Rani & S. Mario (Eds.), *English in the World: global rules, global roles* (pp. 40-50). London: Continuum.
- Seidlhofer, B. (2008). Of norms and mindsets. *Australian Review of Applied Linguistics*, 31(3), 331-337.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.
- Setter, J., Wong, C. S. & Chan, B. H.S. (2010). *Hong Kong English*. Edinburgh: Edinburgh University Press.
- Sewell, A. (2009). World Englishes, English as a lingua franca, and the case of Hong Kong English. *English Today*, 25(1), 37-43.
- Shao, Z. H. (1997). *Yinghan yuyan jiuyu duibi [Researching and contrasting English and Chinese language]*. Shanghai: Huadong ligong daxue



- chubanshe [East China University of Science and Technology Press].
- Sharifian, F. (2003). On cultural conceptualisations. *Journal of Cognition and Culture*, 3(3), 187-207.
- Sharifian, F. (2008). Distributed, emergent cultural cognition, conceptualisation, and language. In R. M. Frank, R. Driven, T. Ziemke & E. Bernandez (Eds.), *Body, language, and mind: Sociocultural situatedness* (pp.109-136). Berlin: Mouton de Gruyter.
- Sharifian, F. (2011). *Cultural conceptualisations and language: Theoretical framework and applications*. Amsterdam: John Benjamins.
- Sharifian, F. (2013). Globalisation and developing metacultural competence in learning English as an International Language. *Multilingual Education*, 3(7). Retrieved from www.multilingual-education.com/content/3/1/7. Accessed 10 July 2014.
- Sinclair, J. (1966). Beginning the study of lexis. In C.E. Bazell, J. C. Catford, M. A. K. Halliday & R.H. Robins (Eds.), *In memory of J.R. Firth* (pp. 410-430). London: Longman.
- Sinclair, J. (1987). Collocation: a progress report. R. Steele & T. Threadgold (Eds.), *Language topics: essays in honour of Michael Halliday* (pp. 319-331). Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, 9(1), 75-106.
- Sinclair, J. (1998). The lexical item. In W. Weigand (Ed.), *Contrastive lexical semantics* (pp. 1-24). Amsterdam: John Benjamins.
- Sinclair, J. (2003). *Reading concordances: an introduction*. Harlow: Pearson Longman.
- Sinclair, J. (2004). *Trust the text: language, corpus and discourse*. London: Routledge.
- Skandera, P. (2007). *Phraseology and culture in English*. Berlin: Mouton de Gruyter.
- Smith, L. E. (Ed.). (1981). *English for cross-cultural communication*. London: Macmillan.
- Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 75-90), Chicago: Illinois University Press.
- Spolsky, B. (1989). *Conditions for second language learning: introduction to a general theory*. Oxford: Oxford University Press.
- Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
- Strevens, P. (1980). *Teaching English as an international language*. Oxford: Pergamon.

- Strevens, P. (1985). Standards and the standard language. *English Today*, (1/2), 5-8.
- Stubbs, M. (1996). *Texts and corpus analysis*. Oxford: Blackwell.
- Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Stubbs, M. (2009). The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics*, 30(1), 115-137.
- Sun, L. (1989). Yingyu guobie bianti de yanjiu he yingyuzaizhongguo [Research on English varieties and English in China]. *Waiguo Yu [Foreign Languages]*, 60(2), 17-23.
- Sun, L. (2011). *Degree adverbs in Hong Kong and Singapore English: a corpus-based investigation*. PhD Dissertation, the University of Hong Kong.
- Svartvik, J., & Leech, G. N. (2006). *English: one tongue, many voices*. Basingstoke: Palgrave Macmillan.
- Swain, M., & Deters, P. (2007). "New" mainstream SLA theory: expanded and enriched. *The Modern Language Journal*, 91(s1), 820-836.
- Sweet, H. (1910). *The sounds of English: an introduction to phonetics*. Oxford: Clarendon Press.
- Tai, J. H-Y. (1985). Temporal sequences and Chinese word order. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 49-72). Amsterdam: John Benjamins.
- Tarone, E. (2007). Sociolinguistic approaches to second language acquisition research 1997-2007. *The Modern Language Journal*, 91(S1), 837-848.
- Taylor, R. L. & Legge, J. (2011). *Confucius, the Analects: The Path of the Sage: Selections Annotated & Explained*. Woodstock, Vt: SkyLight Paths Publishing.
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10(1), 1-13.
- Teubert, W., & Čermáková, A. (2007). *Corpus linguistics: a short introduction*. London: Continuum.
- The World in 2013: ICT Facts and Figures (2014). Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>. Accessed 2 May 2014.
- Thomas, J., & Short, M. (1996). *Using corpora for language research: studies in honour of Geoffrey Leech*. London: Longman.
- Timmis, I. (2002). Native-speaker norms and international English: a classroom view. *ELT Journal*, 56(3), 240-249.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Trudgill, P. (1984). *Language in the British Isles*. Cambridge: Cambridge University Press.



- Trudgill, P. (1999). Standard English: what it isn't. In T. Bex & R. Watts (Eds.), *Standard English: the widening debate* (pp.117-128). London: Routledge.
- Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.
- Trudgill, P., & Hannah, J. (2008). *International English: a guide to varieties of standard English* (5th edn.). London: Hodder Education.
- Wang, L. (1992). *Zhongguo xiandai yufa [Grammar of Chinese language]*. Shanghai: Shanghai Shudian [Shanghai Press].
- Wang, Q. (2007). The national curriculum changes and their effects on English language teaching in the People's Republic of China. *International Handbook of English Language Teaching*, 15(1), 87-105.
- Wang, R. P. (1991). Zhongguo yingyu shi keguan cunzai [China English is an objective reality]. *Jiefangjun waiyu xueyuan xuebao [Journal of PLA Foreign Languages Institute]*, 1, 1-8.
- Wang, R. P. (1994). *Shuo dong dao xi huayingyu [Talking about English]*. Beijing: Waiyu jiaoxue yu yanjiu chubanshe [Foreign language teaching and research press].
- Wang, W. B. & Ma, D. (2002). Zhongguo yingyu ji qi biaoda xingshi [Chinese English and its expressions]. *Dalian minzu xueyuan xuebao [Journal of Dalian Nationalities University]*, 14, 55-8.
- Wang, Z. Y. (2006). *Yinghan yuwen wenti mianmianguan [Aspects on English-Chinese contrastive research]*. Beijing: Gaodeng jiaoyu chubanshe [Higher education press].
- Ward, I. C. (1929). *The phonetics of English*. Cambridge: Heffer.
- Wei, D. Y. (2007). *A Chinese-English Dictionary*. Beijing: Foreign Language Teaching and Research Press.
- Wei, R. N., & Su, J. Z. (2012). The statistics of English in China. *English Today*, 28(3), 10-14.
- Wen, Q. F. (2012). Teaching English as an International Language in mainland China. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: implications for language education* (pp.79-93). Dordrecht: Springer.
- West, C. (1992). A matter of life and death. *October--The MIT Press*, 61, 20-23. Retrieved from <http://www.mitpressjournals.org/loi/octo>. Accessed 2 May 2013.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Widdowson, H. (1997). EIL, ESL, EFL: Global issues and local interests. *World Englishes*, 16(1), 135-146.
- Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
- Wilkes, G. A. (1978). *A dictionary of Australian colloquialisms*. Fontana:



- Collins.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied linguistics*, 21(4), 463-489.
- Wu, G. H. (2010). *Hanying da cidian [Chinese-English dictionary]*. Shanghai: shanghai yiwen chubanshe [Shanghai Translation Publishing House].
- Wulff, S. (2013). *Words and idioms: The Oxford handbook of construction grammar*. Oxford: Oxford University Press.
- Xie, Y. J. (1990). *Xiandai hanyu ouhua yufa gailun [Introduction to Europeanisation of Chinese]*. Hong Kong: Guangming Book Company.
- Xie, Z. J. (1995). Zhongguoyingyu: kuawenhua jiaojizhong de ganraobianti [China English: Interference variety in cross-cultural communication]. *Xiandai Waiyu [Modern Languages]*, 70(4), 7-11.
- Xu, Z. C. (2010). *Chinese English. Features and implications*. Hong Kong: Open University of Hong Kong Press.
- Xu, Z. C. (2013). Globalisation, culture and ELT materials: a focus on China. *Multilingual Education*, 3(6). Retrieved from www.multilingual-education.com/content/3/1/6. Accessed 10 July 2014.
- Yang, H. Z. (1983). JDEST corpus [Electronic data].
- Yang, H. Z. (1998). Daxue yingyu siliuji kaoshi huigu yu zhanwang [Reflection on the CET4 and CET6]. *Waiyujie [Foreign Language World]*, 03, 28-32.
- Yang, H. Z. (2000). Cong siliuji kaoshi kan woguo daxue shengdang qiande shijiyingyu nengli [On the English proficiency of college students from CET4 and CET6]. *Waiyujie [Foreign Language World]*, 77(1), 46-52.
- Yang, H. Z. (2003). Daxue yingyu siliuji kaoshi shi wunian huigu [Reflection on the fifteen years of CET4 and CET6]. *Waiguoyu [Journal of foreign languages]*, 145(3), 21-29.
- Yang, J. (2005). Lexical innovations in China English. *WorldEnglishes*, 24(4), 425-436.
- Yang, J. (2006). Learners and users of English in China. *English Today*, 22(1), 3-10.
- Yang, J. (2009). Chinese borrowings in English. *WorldEnglishes*, 28(1), 90-106.
- Young, M. Y. (2006). Macao students' attitudes toward English: a post-1999 survey. *World Englishes*, 25(3-4), 479-490.
- Yu, X. (2009). *Nativized lexical use in China's English newspapers*. PhD dissertation, Najing University, China.
- Yuan, B. & Church, S. K. (2006). *Oxford Beginner's Chinese*

Dictionary. Oxford: Oxford University Press.

- Zeng, F. & Lu, D. (2006). Dui zhongguoyingyubiantianjiuzhongyouguan fang xiangxing wen ti de sikao [Towards the directions of China English as a variety]. *Huizhou xueyuan xuebao [Journal of Huizhou University]*, 26(2), 51-55.
- Zhang, P. C. (1995). Shiyongmudiyuguobiebian-ti-ye tan zhongguo yingyu [The international or intranational use of English and its national varieties]. *Xiandai Waiyu [Modern Languages]*, 69(03), 16-21.
- Zhang, W. (2012). Chinese-English code-mixing among China's netizens. *English Today*, 28(1), 40-52.
- Zhuang, Y. C. (2000). Ye tan zhongshiyingyu [Guard against Chinglish]. *Zhongguo Fanyi [Chinese Translators Journal]*, 6, 7-10.
- Zipf, G. K. (1935). *The psycho-biology of language*. Boston: Houghton-Mifflin.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Boston: Addison-Wesley Press.



Appendix 1: THE BNC BASIC TAGSET

Tag	Description
<u>AJ0</u>	Adjective (general or positive) (e.g. <i>good, old, beautiful</i>)
<u>AJC</u>	Comparative adjective (e.g. <i>better, older</i>)
<u>AJS</u>	Superlative adjective (e.g. <i>best, oldest</i>)
<u>AT0</u>	Article (e.g. <i>the, a, an, no</i>)
<u>AV0</u>	General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. <i>often, well, longer (adv.), furthest</i>).
<u>AVP</u>	Adverb particle (e.g. <i>up, off, out</i>)
<u>AVQ</u>	Wh-adverb (e.g. <i>when, where, how, why, wherever</i>)
<u>CJC</u>	Coordinating conjunction (e.g. <i>and, or, but</i>)
<u>CJS</u>	Subordinating conjunction (e.g. <i>although, when</i>)
<u>CJT</u>	The subordinating conjunction <i>that</i>
<u>CRD</u>	Cardinal number (e.g. <i>one, 3, fifty-five, 3609</i>)
<u>DPS</u>	Possessive determiner-pronoun (e.g. <i>your, their, his</i>)
<u>DT0</u>	General determiner-pronoun: i.e. a determiner-pronoun which is not a DTQ or an AT0.
<u>DTQ</u>	Wh-determiner-pronoun (e.g. <i>which, what, whose, whichever</i>)
<u>EX0</u>	Existential there, i.e. <i>there</i> occurring in the <i>there is ...</i> or <i>there are ...</i> construction
<u>ITJ</u>	Interjection or other isolate (e.g. <i>oh, yes, mhm, wow</i>)
<u>NN0</u>	Common noun, neutral for number (e.g. <i>aircraft, data, committee</i>)
<u>NN1</u>	Singular common noun (e.g. <i>pencil, goose, time, revelation</i>)
<u>NN2</u>	Plural common noun (e.g. <i>pencils, geese, times, revelations</i>)
<u>NP0</u>	Proper noun (e.g. <i>London, Michael, Mars, IBM</i>)
<u>ORD</u>	Ordinal numeral (e.g. <i>first, sixth, 77th, last</i>) .
<u>PNI</u>	Indefinite pronoun (e.g. <i>none, everything, one [as pronoun], nobody</i>)
<u>PNP</u>	Personal pronoun (e.g. <i>I, you, them, ours</i>)
<u>PNQ</u>	Wh-pronoun (e.g. <i>who, whoever, whom</i>)
<u>PNX</u>	Reflexive pronoun (e.g. <i>myself, yourself, itself, ourselves</i>)
<u>POS</u>	The possessive or genitive marker 's or '
<u>PRF</u>	The preposition <i>of</i>
<u>PRP</u>	Preposition (except for <i>of</i>) (e.g. <i>about, at, in, on, on behalf of, with</i>)
<u>PUL</u>	Punctuation: left bracket - i.e. (or [
<u>PUN</u>	Punctuation: general separating mark - i.e. . , ! , : ; - or ?
<u>PUQ</u>	Punctuation: quotation mark - i.e. 'or ''
<u>PUR</u>	Punctuation: right bracket - i.e.) or]
<u>TO0</u>	Infinitive marker <i>to</i>

<u>UNC</u>	Unclassified items which are not appropriately considered as items of the English lexicon.
<u>VBB</u>	The present tense forms of the verb BE, except for <i>is</i> , 's: i.e. <i>am</i> , <i>are</i> , 'm, 're and <i>be</i> [subjunctive or imperative]
<u>VBD</u>	The past tense forms of the verb BE: <i>was</i> and <i>were</i>
<u>VBG</u>	The -ing form of the verb BE: <i>being</i>
<u>VBI</u>	The infinitive form of the verb BE: <i>be</i>
<u>VBN</u>	The past participle form of the verb BE: <i>been</i>
<u>VBZ</u>	The -s form of the verb BE: <i>is</i> , 's
<u>VDB</u>	The finite base form of the verb BE: <i>do</i>
<u>VDD</u>	The past tense form of the verb DO: <i>did</i>
<u>VDG</u>	The -ing form of the verb DO: <i>doing</i>
<u>VDI</u>	The infinitive form of the verb DO: <i>do</i>
<u>VDN</u>	The past participle form of the verb DO: <i>done</i>
<u>VDZ</u>	The -s form of the verb DO: <i>does</i> , 's
<u>VHB</u>	The finite base form of the verb HAVE: <i>have</i> , 've
<u>VHD</u>	The past tense form of the verb HAVE: <i>had</i> , 'd
<u>VHG</u>	The -ing form of the verb HAVE: <i>having</i>
<u>VHI</u>	The infinitive form of the verb HAVE: <i>have</i>
<u>VHN</u>	The past participle form of the verb HAVE: <i>had</i>
<u>VHZ</u>	The -s form of the verb HAVE: <i>has</i> , 's
<u>VM0</u>	Modal auxiliary verb (e.g. <i>will</i> , <i>would</i> , <i>can</i> , <i>could</i> , 'll, 'd)
<u>VVB</u>	The finite base form of lexical verbs (e.g. <i>forget</i> , <i>send</i> , <i>live</i> , <i>return</i>) [Including the imperative and present subjunctive]
<u>VVD</u>	The past tense form of lexical verbs (e.g. <i>forgot</i> , <i>sent</i> , <i>lived</i> , <i>returned</i>)
<u>VVG</u>	The -ing form of lexical verbs (e.g. <i>forgetting</i> , <i>sending</i> , <i>living</i> , <i>returning</i>)
<u>VVI</u>	The infinitive form of lexical verbs (e.g. <i>forget</i> , <i>send</i> , <i>live</i> , <i>return</i>)
<u>VVN</u>	The past participle form of lexical verbs (e.g. <i>forgotten</i> , <i>sent</i> , <i>lived</i> , <i>returned</i>)
<u>VVZ</u>	The -s form of lexical verbs (e.g. <i>forgets</i> , <i>sends</i> , <i>lives</i> , <i>returns</i>)
<u>XX0</u>	The negative particle <i>not</i> or <i>n't</i>
<u>ZZ0</u>	Alphabetical symbols (e.g. <i>A</i> , <i>a</i> , <i>B</i> , <i>b</i> , <i>c</i> , <i>d</i>)

Appendix 2: Top 50 adjectives in L-1 position (in order of frequency)

AREA			CHILD		
CEC		BNC	CEC		BNC
1.	rural	rural	1.	young	young
2.	urban	other	2.	other	other
3.	coastal	urban	3.	little	small
4.	other	large	4.	Chinese	only
5.	stricken	local	5.	own	handicapped
6.	mountainous	particular	6.	disabled	deaf
7.	total	different	7.	many	early
8.	scenic	small	8.	small	black
9.	poor	residential	9.	early	little
10.	key	new	10.	second	sick
11.	autonomous	main	11.	only	dependent
12.	different	wide	12.	age	national
13.	residential	certain	13.	first	unborn
14.	inhabited	whole	14.	old	poor
15.	many	geographical	15.	poor	individual
16.	administrative	specific	16.	third	local
17.	pastoral	key	17.	sick	pre-school
18.	western	surrounding	18.	Japanese	illegitimate
19.	suburban	important	19.	female	disabled
20.	large	sensitive	20.	rural	new
21.	military	industrial	21.	local	surviving
22.	new	remote	22.	unborn	good
23.	functional	central	23.	mere	normal
24.	sown	inner-city	24.	good	working-class
25.	certain	grey	25.	noisy	old
26.	whole	major	26.	retarded	teenage
27.	central	metropolitan	27.	speaking	British
28.	such	built-up	28.	American	happy
29.	green	vast	29.	autistic	Asian
30.	specific	conservation	30.	blind	hearing
31.	remote	coastal	31.	future	healthy
32.	producing	total	32.	handicapped	particular
33.	vast	following	33.	illegitimate	autistic
34.	following	difficult	34.	innocent	starving
35.	agricultural	deprived	35.	late	innocent
36.	various	extensive	36.	several	dead
37.	small	designated	37.	dead	naughty
38.	hit	open	38.	little	lost
39.	local	broad	39.	born	middle-class
40.	eastern	only	40.	foreign	disturbed
41.	underdeveloped	various	41.	fourth	special
42.	arid	affected	42.	happy	bilingual
43.	major	assisted	43.	lovely	adopted
44.	northern	immediate	44.	newborn	deprived
45.	subject	protected	45.	russian	grown-up
46.	same	wooded	46.	spoilt	ill
47.	controlled	depressed	47.	traditional	gifted
48.	commercial	inner	48.	bad	Indian
49.	affected	huge	49.	few	white
50.	important	disadvantaged	50.	homeless	modern



(Continued).

COUNTRY	
CEC	BNC
1. other	other
2. developed	European
3. Asian	developing
4. foreign	developed
5. different	different
6. western	foreign
7. whole	African
8. beautiful	western
9. European	whole
10. many	open
11. own	American
12. speaking	industrialized
13. African	surrounding
14. socialist	low
15. various	industrial
16. American	poor
17. neighboring	small
18. new	industrialised
19. relevant	rich
20. third	neighbouring
21. big	Asian
22. income	individual
23. first	beautiful
24. major	new
25. certain	only
26. entire	particular
27. great	communist
28. poor	various
29. populous	black
30. advanced	English
31. many	old
32. recipient	Arab
33. large	Basque
34. open	advanced
35. few	big
36. several	socialist
37. affected	certain
38. agricultural	entire
39. powerful	native
40. rich	Mediterranean
41. small	great
42. Arab	independent
43. individual	single
44. fellow	wild
45. industrialized	democratic
46. only	eastern
47. sovereign	large
48. respective	tropical
49. such	British
50. vast	major

EYE	
CEC	BNC
1. own	blue
2. blue	brown
3. black	grey
4. green	black
5. large	green
6. big	dark
7. brown	bright
8. beady	closed
9. small	narrowed
10. childlike	pale
11. very	large
12. red	yellow
13. bright	very
14. sharp	red
15. watchful	big
16. closed	beautiful
17. filled	little
18. gloomy	open
19. human	wide
20. narrow	golden
21. dark	sharp
22. many	new
23. same	bloodshot
24. sleepy	ice-blue
25. angry	small
26. Chinese	watchful
27. clear	beady
28. new	hooded
29. set	keen
30. tired	tired
31. American	dead
32. beautiful	sightless
33. blazing	sore
34. deep	cold
35. expressive	startled
36. open	other
37. swollen	red-rimmed
38. expectant	half-closed
39. intense	frightened
40. listless	glittering
41. lustreless	hard
42. mammalian	huge
43. pretty	blind
44. puffy	western
45. round	human
46. shifty	heavy-lidded
47. astonished	angry
48. colored	sad
49. curious	round
50. empty	watery



(Continued).

FACT		
CEC		BNC
1.	important	very
2.	other	mere
3.	cultural	simple
4.	social	actual
5.	many	important
6.	key	sad
7.	external	historical
8.	main	obvious
9.	major	interesting
10.	contextual	well-known
11.	various	basic
12.	cognitive	hard
13.	linguistic	curious
14.	decisive	central
15.	historical	significant
16.	affective	undeniable
17.	such	scientific
18.	following	social
19.	basic	inescapable
20.	pragmatic	jurisdictional
21.	essential	objective
22.	internal	remarkable
23.	crucial	new
24.	total	physical
25.	environmental	accomplished
26.	different	accepted
27.	influential	psychological
28.	several	empirical
29.	soluble	indisputable
30.	climatic	vital
31.	economic	bare
32.	psychological	established
33.	significant	fundamental
34.	relevant	known
35.	new	proven
36.	semantic	striking
37.	negative	surprising
38.	only	key
39.	situational	stark
40.	second	strange
41.	above	unfortunate
42.	actual	harsh
43.	affecting	undoubted
44.	objective	incontrovertible
45.	fundamental	observable
46.	random	relevant
47.	simple	astonishing
48.	institutional	only
49.	nuclear	certain
50.	political	disturbing

HOME		
CEC		BNC
1.	own	new
2.	new	residential
3.	old	private
4.	back	matrimonial
5.	ancestral	ideal
6.	former	stately
7.	Chinese	old
8.	second	good
9.	native	Scottish
10.	ideal	other
11.	spiritual	local
12.	straight	permanent
13.	original	foster
14.	rural	British
15.	real	comfortable
16.	small	happy
17.	army	long
18.	first	spiritual
19.	many	ancestral
20.	ancient	young
21.	comfortable	natural
22.	humble	broken
23.	native	national
24.	respective	parental
25.	online	real
26.	only	small
27.	beautiful	temporary
28.	different	traditional
29.	distant	only
30.	famous	little
31.	floral	independent
32.	permanent	original
33.	pleasant	safe
34.	style	individual
35.	symbolic	previous
36.	temporary	nearby
37.	traditional	common
38.	underground	large
39.	various	lovely
40.	very	available
41.	American	single
42.	cellular	existing
43.	common	unbeaten
44.	elderly	empty
45.	enthusiastic	proper
46.	eternal	repossessed
47.	familiar	secure
48.	main	modern
49.	quiet	nice
50.	unique	suitable



(Continued).

MAN		
CEC		BNC
1.	old	young
2.	young	old
3.	Chinese	other
4.	other	dead
5.	good	good
6.	superior	little
7.	rich	big
8.	great	white
9.	dark	poor
10.	dead	great
11.	white	working
12.	bad	new
13.	many	black
14.	American	married
15.	wise	gay
16.	big	rich
17.	real	local
18.	sick	wise
19.	few	tall
20.	black	elderly
21.	modern	middle-aged
22.	aged	only
23.	first	ordinary
24.	honest	different
25.	poor	thin
26.	thin	nice
27.	fat	blind
28.	armed	armed
29.	elderly	powerful
30.	tall	small
31.	brave	strong
32.	looking	honest
33.	ordinary	handsome
34.	only	reasonable
35.	military	single
36.	own	brave
37.	several	modern
38.	strong	sick
39.	certain	strange
40.	mean	homosexual
41.	single	free
42.	common	wealthy
43.	Shuang	real
44.	Mexican	wild
45.	evil	large
46.	powerful	medical
47.	such	holy
48.	young	military
49.	blind	British
50.	perfect	dying

PEOPLE		
CEC		BNC
1.	young	other
2.	other	young
3.	Chinese	elderly
4.	own	old
5.	disabled	local
6.	many	disabled
7.	only	black
8.	school-age	ordinary
9.	first	deaf
10.	old	British
11.	poor	different
12.	sick	unemployed
13.	Japanese	poor
14.	rural	only
15.	local	new
16.	unborn	real
17.	good	white
18.	noisy	indigenous
19.	retarded	homeless
20.	American	common
21.	autistic	certain
22.	blind	good
23.	future	single
24.	handicapped	ill
25.	illegitimate	American
26.	innocent	professional
27.	several	important
28.	dead	nice
29.	little	innocent
30.	born	rich
31.	foreign	retired
32.	fourth	English
33.	happy	little
34.	lovely	tribal
35.	newborn	decent
36.	Russian	various
37.	spoilt	working-class
38.	traditional	Irish
39.	bad	Jewish
40.	homeless	famous
41.	lovable	Chinese
42.	school	handicapped
43.	white	normal
44.	affected	native
45.	aged	Scottish
46.	black	blind
47.	disobedient	religious
48.	hungry	working
49.	joyful	intelligent
50.	bilingual	interesting



(Continued).

PROBLEM		
CEC		BNC
1.	serious	major
2.	major	social
3.	environmental	serious
4.	many	main
5.	social	real
6.	difficult	particular
7.	big	other
8.	practical	economic
9.	such	financial
10.	main	only
11.	same	similar
12.	new	environmental
13.	key	technical
14.	other	practical
15.	financial	difficult
16.	technical	new
17.	common	political
18.	similar	common
19.	economic	specific
20.	real	fundamental
21.	potential	legal
22.	own	special
23.	prominent	big
24.	structural	different
25.	basic	basic
26.	first	personal
27.	outstanding	immediate
28.	above	potential
29.	fundamental	great
30.	linguistic	central
31.	important	emotional
32.	specific	severe
33.	technological	behavioural
34.	following	psychological
35.	unsolved	growing
36.	related	current
37.	severe	medical
38.	central	considerable
39.	certain	additional
40.	complex	complex
41.	current	methodological
42.	mental	sexual
43.	second	structural
44.	various	intractable
45.	ecological	marital
46.	global	important
47.	urgent	administrative
48.	seated	obvious
49.	immediate	old
50.	mentioned	significant

QUESTION		
CEC		BNC
1.	following	important
2.	first	difficult
3.	many	other
4.	such	following
5.	second	fundamental
6.	same	key
7.	simple	simple
8.	no	open
9.	open	whole
10.	genuine	interesting
11.	provocative	crucial
12.	few	specific
13.	important	basic
14.	further	big
15.	above	central
16.	crucial	awkward
17.	interesting	serious
18.	other	only
19.	ended	parliamentary
20.	specific	unanswered
21.	choice	obvious
22.	experiment	Jewish
23.	general	real
24.	key	main
25.	old	political
26.	fundamental	direct
27.	last	general
28.	Taiwan	good
29.	big	rhetorical
30.	major	detailed
31.	next	relevant
32.	real	new
33.	several	critical
34.	third	vexed
35.	whole	wrong
36.	main	national
37.	unanswered	major
38.	basic	particular
39.	different	different
40.	difficult	similar
41.	new	silly
42.	central	vital
43.	great	personal
44.	interrogative	certain
45.	serious	postal
46.	standard	closed
47.	such	technical
48.	urgent	leading
49.	common	pertinent
50.	particular	supplementary



(Continued).

SYSTEM		
CEC		BNC
1.	conceptual	political
2.	legal	new
3.	economic	nervous
4.	financial	open
5.	social	legal
6.	political	social
7.	linguistic	economic
8.	educational	immune
9.	new	educational
10.	ecological	electoral
11.	immune	monetary
12.	semantic	financial
13.	administrative	existing
14.	based	whole
15.	nervous	other
16.	whole	multiparty
17.	cognitive	old
18.	solar	penal
19.	complete	solar
20.	socialist	different
21.	cultural	judicial
22.	complex	current
23.	corporate	complex
24.	binary	visual
25.	current	national
26.	sound	British
27.	regulatory	democratic
28.	coherent	digestive
29.	democratic	comprehensive
30.	judicial	integrated
31.	different	formal
32.	grammatical	two-tier
33.	old	regulatory
34.	theoretical	international
35.	industrial	administrative
36.	other	parliamentary
37.	monetary	interactive
38.	visual	similar
39.	dynamic	parallel
40.	international	closed
41.	medical	complete
42.	Chinese	entire
43.	lexical	effective
44.	literary	distributed
45.	theatrical	large
46.	comprehensive	computerised
47.	feudal	electronic
48.	organizational	traditional
49.	own	federal
50.	limited	particular

THING		
CEC		BNC
1.	same	other
2.	other	only
3.	such	whole
4.	only	good
5.	first	important
6.	important	real
7.	many	bad
8.	good	little
9.	new	different
10.	whole	great
11.	different	terrible
12.	bad	main
13.	right	certain
14.	real	funny
15.	strange	strange
16.	little	new
17.	few	old
18.	old	interesting
19.	great	wrong
20.	terrible	nice
21.	last	wonderful
22.	beautiful	poor
23.	precious	small
24.	concrete	natural
25.	abstract	big
26.	next	living
27.	certain	beautiful
28.	difficult	sensible
29.	big	odd
30.	earthly	simple
31.	various	difficult
32.	interesting	very
33.	main	damn
34.	nice	stupid
35.	proper	various
36.	small	silly
37.	natural	awful
38.	familiar	single
39.	unpleasant	obvious
40.	common	dangerous
41.	evil	exciting
42.	objective	damned
43.	other	extraordinary
44.	wonderful	bloody
45.	own	personal
46.	stupid	useful
47.	trivial	remarkable
48.	valuable	similar
49.	fine	dreadful
50.	mean	ordinary



(Continued)

WAY		
CEC		BNC
1.	same	other
2.	different	different
3.	long	long
4.	effective	only
5.	other	new
6.	only	various
7.	many	effective
8.	new	similar
9.	own	wrong
10.	round	good
11.	various	possible
12.	traditional	easy
13.	good	particular
14.	following	normal
15.	Chinese	usual
16.	important	certain
17.	systematic	old
18.	efficient	following
19.	similar	simple
20.	particular	alternative
21.	certain	obvious
22.	unique	efficient
23.	appropriate	traditional
24.	comprehensive	practical
25.	right	hard
26.	big	useful
27.	several	important
28.	direct	small
29.	proper	appropriate
30.	indirect	main
31.	special	big
32.	natural	separate
33.	possible	strange
34.	specific	ideal
35.	old	whole
36.	orderly	bad
37.	alternative	natural
38.	wrong	special
39.	common	systematic
40.	convenient	correct
41.	half	convenient
42.	scientific	positive
43.	basic	proper
44.	complex	general
45.	economical	direct
46.	correct	ordinary
47.	easy	excellent
48.	simple	subtle
49.	main	American
50.	subtle	sensible





**The Hong Kong
Institute of Education Library**

For private study or research only.
Not for publication or further reproduction.