A comparative investigation of collocation, colligation and semantic association in

English and Chinese: a test of Chinese on the applicability of three major concepts as

defined in Lexical Priming Theory

A Thesis Submitted in accordance with the requirements of

The University of Liverpool and The Department of English at Xi'an Jiaotong-

Liverpool University

for the Degree of Doctor of Philosophy

By WANG Luojia

200858962

Nov. 2018

ABSTRACT

It has been more than ten years since the official 2005 publication of the Lexical Priming Thoery (LPT), when various applications and expansions have been made to advance this theory in many dimensions. The majority of this kind of study has been done in English, while little research has been done in Mandarin Chinese—a genetically different language from English, with a notable exception of Hoey and Shao (2015), who addressed issue by undertaking a preliminary study of several notions as defined in LPT.

In order to fill in this gap, this thesis explores the fundamental concepts noted in LPT, namely collocation, colligation and semantic association, to provide more evidence of LPT's ability to explain the linguistic features of Chinese and, based on that evidence, to determine whether LPT can be further advanced or expanded to address the characteristics of Chinese. Furthermore, we attempt to expand LPT's dimensions in the field by comparing the lexical and syntactical similarities in and differences between English and Chinese. By examining the evidence presented by the given nodes (*world* and *世界 shi4jie4*) and the clusters containing the nodes, we aim to assess the potential of applying LPT in cross-linguistic studies of English and Chinese.

A holistic analysis of the nodes and phraseology units containing the nodes show that the three concepts as defined in LPT are also applicable to Chinese. With respect of collocation, we have found that collocation of Mandarin Chinese operates on both single characters and nestings. With respect of colligation, we have found that there is an interaction between the node items' inter- and intra-colligation. The colligation is primed in an interaction with collocation and semantic association. The parameters interchange with and are dependent on each other.

The interaction between collocation, colligation and semantic association operates on *世界 shi4jie4* implies that the relation between grammar and lexis of Chinese may need to be re-considered as well. First, the "borrowed" grammatical class of "word" needs to be re-defined, especially in a cross-linguistic study between English and Chinese. The English definition of word are not tied in with Chinese due to the

latter's distinctive morphological features. We propose, with evidence found for 世界 *shi4jie4* that *nesting* as defined in LPT could be applied as a basic corresponding pattern and departure point for cross-linguistic study between English and Chinese. Second, we propose a new approach of describing Chinese grammar after testing the validity of LPT for Mandarin Chinese that the grammatical category a Chinese character and nesting is assigned shall be an outcome of the character or nesting's grammatical primings. The grammar of Chinese is an accumulation and interweaving of the primings of a character or a nesting's collocation, colligation and semantic association. Components of nestings are primed to combine in certain way (order), to occur at certain position within a nominal group, and to serve a certain function in a sentence.

By comparing the differences and similarities exhibited by cross-linguistic equivalence (*world* and 世界 *shi4jie4*), we discovered that there is no absolute "similarity" or "differences" regarding the lexical and syntactical priming features between the two languages. The divergence lies in the congruence. We could not restrict the cross-linguistic study to one unit, rather, we ought to analyse the features from the perspective of co-textual interaction. A more abstract unit, for example, the semantic association needs to be considered.

Key Words: The Lexical Priming Theory; Collocation; Colligation; Semantic Association; Cross-linguistic study

ACKNOWLEDGEMENTS

Too many people deserve my sincere thanks in the process of completing this thesis. First and foremost, I wish to express my deepest thanks and gratitude to my supervisor Professor Michael Hoey, who had provided unreserved professional guidance for the writing, especially during his retirement. Hundreds of silly mistakes were made in the first PhD year. My supervisor, with his great patience, assisted me to correct those mistakes and recommended useful books to me. For articles which were hard to get a copy in China, Michael went blocks away to scan pages which he thought would be helpful for me. During the four years of being supervised by Michael, I took him not only as a teacher, but also a close friend.

My thanks also go to my based Supervisor at Xi'an Jiaotong-Liverpool University (XJTLU), Professor Zhoulin Ruan, who has been so nice to provide me with comments and suggestions, especially in the course of viva preparation and thesis revision. The thesis would not be completed without his great support and kindness. I'm grateful to Professor Ruan for treating me throughout with understanding, patience and valuable suggestions.

I'd like, take this opportunity to express my great thanks to my family. My husband and my parents not only provide financial support to me, but also try their best to establish a "utopian" study environment for me. They don't let me get involved in any family chores which could waste my time and labour. I'd like, specifically, express my gratitude to the most valuable gift I've ever had—my son. His birth encourages me to work hard so as to set a good example for his study life in the

future.

The same gratitude also goes to my universities: The University of Liverpool, Xi'an Jiaotong-Liverpool University, and my working place Dalian University of Foreign Languages. The generosity has been matched by the faculty's assistance in the process of viva prearation and thesis submission. With respect to colleagues at my university, I would like to thank them for covering my teaching job in my absence. Taking over my job without showing a hint of disgruntlement has been much appreciated and was entirely typical of them.

TABLE OF CONTENTS

List of Abbreviation

B.C.	Before Computers
BNC	British National Corpus
BoE	Collins Bank of English
CLEC	Chinese Learner English Corpus
CNC	Chinese National Corpus
DW	Database of *wolrd*/世界 *shi4jie4*
ELU	Extended Lexical Units
FLOB	Freiburg-LOB corpus of British English
JDEST	Jiao Da English Corpus for Science and Technology
L2	Second Language
LCMC	Lancaster Corpus of Mandarin Chinese
LDN-H	London Borough of Hackney
LPT	Lexical Priming Theory
MC	Mutual Correspondence
SCO	Spoken Liverpool English
SJTU	Shanghai Jiao Tong University Parallel Corpus

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Inspiration for writing this thesis

When I was a college student at the English Department of Dalian University of Foreign Languages (located in Dalian City, China) ten years ago, I greatly respected a boy who had been nicknamed "the Prince of Grammar" by our classmates. He was popular with both teachers and students because he always got full marks on his English grammar examinations.

He was a nice person and always ready to help others. So I, like other classmates, often came to him for grammar help. However, I was sometimes not satisfied with his answers to my questions. When questions concerning collocations were asked, such as why the collocation of *learn the knowledge* is not as appropriate as *gain/acquire the knowledge*, he would answer perfunctorily: "You need to read more authentic materials and improve your *linguistic sense* (语感 *yu3gan3*[1] in Chinese, referring to one's intuitive understanding of a language); in other words, you lack linguistic wisdom." Lacking linguistic wisdom? Though I secretly agreed with him that being exposed to more authentic materials would greatly help my English

---

[1] The letters following the Chinese characters 语感 are called *Pinyin*—an official Romanisation system for standard Chinese. This system includes four diacritics denoting tones. For typing convenience, the tones are written in numbers (1-4) to represent the four tones ā, á, ǎ and à For example, *yu3* equals to *yǔ*. *Linguistic sense* is a literal translation to the 语感 *yu3gan3* said by the "Prince of Grammar", where the character of 语 *yu3* means *utterance*, and 感 *gan3* means *sense* or *intuition*, depending on the character preceding to the character of 感 *gan3*.

learning, the answer "You need to improve your *yu3gan3*" did not satisfy me.

Ten years later, I am now an English teacher at the university where I used to study. The questions most frequently asked by my students are of the same kind as those I used to ask the Prince of Grammar—such as, "What are the differences between *bulge*, *protrude*, *stick*, *jut,* and *swell*? They all share the sense of *something sticking out from something*, do they also differ regarding grammar?" or, "How does one distinguish *brim* from *rim,* since their Chinese equivalences are the same (边 *bian1*)?" My answer is generally, "You need to improve your *yu3gan3*. After encountering more authentic materials, you will naturally acquire the distinction between *brim* and *rim*." Every time I make this kind of reply to my students, I am reminded of the "Prince of Grammar" from my college years, and realise I owe him an apology for thinking badly of him. He was of course right. *Yu3gan3* plays an important role in answering questions that cannot be explained by grammar solely. But what are reasons for the existence of *yu3gan3*? How does this phenomenon come into being and how can humans establish their *yu3gan3* for first and second language acquisition? Why does a given word have a strong tendency of appearing with another specific word? Why can words not randomly appear in a perfect grammatical pattern to produce a natural meaning? The seed of pursuing the answer to these questions was sown deeply in my heart and inspired me to choose collocation as my starting point for the present study.

1.2 Reconsideration of the relationship between lexis and grammar

Learn the knowledge and *gain the knowledge* follow the same V + N grammatical

structure. However, having been exposed to authentic English materials for years, I tend to feel the former sounds awkward and unnatural, while the latter seems more idiomatic. This sense is also evidenced by the number of concordance lines retrieved from the British National Corpus (BNC), which is a general English corpus containing more than 100 million words. There are 74 concordance lines occurring for the combination of *gain + knowledge,* but no single instance occurring for the combination of *learn + knowledge*. This grammatically perfect example fails to express a native-like meaning.

In a second-language-learning context, grammar always receives extremely close attention and is treated as a systematic rule, while lexis is chosen to drop into grammatical opportunities. This is especially the case in China. China has a long tradition of examination-oriented education, in particular in teaching and learning English. Most English teaching or learning materials in China are designed to familiarise students with specific grammatical constructions, following specific grammatical rules set by the national syllabus and wordlists provided in text books, normally with one or two word-to-word translations provided by dictionary entries. The result is that learners of English in China recognise individual word meanings and produce their utterances according to their Chinese literal translation. For example, it is very common to encounter a L2 learner in China producing a grammatically correct but non-idiomatic sequence, such as *learn the knowledge*. Second language learners memorise and use individual words according to their L1 translation instead of meaningful chunks; or, as Wray (2000:468) puts it,

> knowing which subset of grammatically possible utterances is actually commonly used by native speakers is an immense problem for even the most proficient of non-natives, who are unable to separate out and avoid the grammatical but non-idiomatic sequences.

The separation of grammar and lexis has been termed as "slot and filler" model (Sinclair 1991: 109), wherein sentences and texts are envisaged as slots that need to be filled by a lexicon. The only restraint is grammaticalness, which is thought to be the bone of a language. According to this view, the only thing a second language learner needs do after mastering grammar is to acquire a vocabulary to fill the "slots." However, words do not occur at random in a text, and the "slot and filler" model does not provide enough restraints on consecutive choices to make proper meanings. For example, "learn the knowledge" strictly follows the rule of V + N; however, it is seen as less natural when compared with examples such as *learn English* or *learn to play the piano.*

Sinclair then put forward another model to account for the restraints that are not captured by the "slot and filter" model — "the idiom principle" (ibid: 110), which holds that "a language user has available to him or her a large number of semi-constructed phrases that constitute single choices, even though they might appear to be analysable into segments." This principle can be seen in the simultaneous or intuitive choice of two words, and attempts to combine the potential meaning of the words with the structure demanded by grammar.

The idiom principle inspires us to think about the relationship between grammar and lexis. As Hunston and Francis (1998: 45) stated, "there is a need for grammar that is lexically based." Grammatical structure could be ordered by a specific rule; however, word choice should not be simply taken for granted. They co-select each other to formulate a particular meaning. The view that grammar is systematic and lexis is

loosely organised needs to be reconsidered by taking meaning into consideration.

The idiom principle implies that much of our language is formulaic, or semi-structured. With the development of computer technology and invention of tools for analysing big data, corpus linguistic analyses show, through large collections of texts, that the naturalness of language is realised through considerable use of recurrent patterns of words and constructions. Kjellmer (1987:140) confirmed that "[in] all kinds of texts, collocations are indispensable elements with which our utterances are very largely made. "

Increasing respect for real examples has raised linguists', especially corpus linguists' awareness of the importance of the "naturalness" of language (e.g., Huston and Francis 1998, Sinclair 1991, Stubbs 1996, Tognini-Boneli 2001). Sinclair (1991:6) used a rhetorical sentence to emphasise the importance of studying authentic texts, as opposed to studying linguistics based on writers' intuitions and language prejudices: "[one] does not study all of botany by making artificial flowers." What's more, naturalness "is simply a cover term for the constraints that determine the precise relationship of any fragment of text with the surrounding text" (Sinclair 1991: 6). The crucial word here is "constraints," which is also picked up on by Stubbs (1996:56), who argued that "[speakers] are free, but only within constraints. Individual speakers intend to communicate with one another in the process of the moment to moment interaction. The reproduction of the system is the unintended product of their routine behaviour." Grammar, to a degree, is the major complement contributing to the constraints, and the "slot and filler" relationship may sometimes be broken or reproduced by human beings' internal intention or intuition. Here, we

do not mean we should devalue the role of grammar and support the existence of non-grammatical phrases, but hold the view that grammar and lexis are interchangeable. Studying one while neglecting the other would block our understanding of the whole language. For years, especially when computers were not widely used for language study, traditional grammarians largely relied on introspection and trusted their intuition. According to Sinclair (1991:39), however, the problem with introspection is that "it does not give evidence about usage… Actual usage plays a minor role in one's consciousness of language, and one would be recording largely ideas about language rather than facts of it." We thus suggest that the evidence of both introspection and actual use should be considered in examining a language.

The previous example (*learn the knowledge*) of non-idiomatic meaning seems to indicate there is a strong tendency for sense and syntax to be associated with. While *learn the knowledge* makes little sense, *learn English* and *learn to do something* are more welcomed by advanced learners of English and native speakers alike. Such examples show that collocations are not merely words co-occurring with each other randomly in a context. As Pace-Sigge (2013: 13) put it, "[once] a statistically high frequency of use is established, they can be seen as more than just a chunk of words but rather as meaningful clusters that have 'idiomaticity'. The "idiomaticity" is recycled and stored in the memory and then is subsequently retrieved whenever needed."

Corpus linguists have attempted to define collocation and the relationship between lexis and grammar in several ways, but have not explored why these phenomena

come into being in the first place. There must be a reason why certain words or clusters co-occur, and why they exist in a certain grammatical structure to formulate a certain meaning. The theory of Lexical Priming (2005) suggests why certain words or sequence of words are likely or unlikely to co-occur with other words or word sequences within a particular grammatical pattern to express a particular meaning.

1.3 The Lexical Priming Theory

New ideas need to incubate, in terms of time, foundations, and technologies. Initially accepting collocation as a starting point to describe what Sinclair, Stubbs and others found, Hoey (2005) moved beyond the surface and attempted to account for this phenomenon by adopting Partington's (1998) psychological definition of collocation. Hoey, Sinclair and Mauranen (2006) acknowledged that psycholinguistic experiments could bolster their claims about collocation, but did not conduct any. Wray (2002) was concerned with how the mind worked and highlighted why there were collocations rather than merely co-occurrences of words. She pointed out that collocation occurs as a fluid version of formulaicity, and that formulaic blocks appear as part of first language acquisition. Wary's "formulaic blocks" and Sinclair's "idiomatic principle" suggest linguists have become aware of the psychological dimension of collocation and have brought it into their discussions thereof. Hoey (2005:9) adopted the word "priming" to account for this psychological existence of the linguistic phenomenon, explaining it as follows:

> Priming need[s] not be a permanent feature of the word or word sequence; in principle, indeed, it never is. Every time we use a word, and every time we encounter it anew, the experience either reinforces the priming by confirming an existing association between the word and its co-texts and context, or it weakens the priming, if the encounter introduces the word in an unfamiliar context or

co-text or if we have chosen in our own use of it to override its current priming.

Starting with the pervasiveness of collocation, Hoey further proposed two notions—semantic association and colligation—to integrate lexis, grammar and meaning. Hoey was not the first to propose these notions; before him, Sinclair (1991, 2004a) had used the term "semantic preference" to refer to the meaningful outcome of the complexity of collocational choices, while Halliday (1957) mentioned that a word tends to prefer to occur at a particular position in a clause or sentence (i.e., colligation). Hoey explored and extended Halliday's notion of colligation to text (textual colligation), and used lexical priming to "approach the acquisition, understanding and production of language (either spoken or written) from lexis-driven, not a grammar-driven stance" (Pace-Sigge 2017: xii). Hoey referred to semantic association (preference) as a sum of meaning stored in the language user's mind by a set of semantic members. He accounted for colligation with priming, stating that every word is primed to prefer (or avoid) certain grammatical positions and functions.

Lexical Priming Theory (LPT) has shed new light on the relationship between lexis and grammar, from a psychological perspective. It is intended as a bridge between "externalized language" and "internalised language" (Chomsky 1986).

Since first published in 2005, Lexical Priming (LP) has been largely tested on in the English language (e.g., Pace-Sigge 2013; Patterson 2014). Little research, however, has been done involving such non-European languages as Mandarin Chinese, a language that is distinct from English and other Indo-European languages from many perspectives. Mandarin Chinese is a *lingua franca* used by millions of speakers, both

in mainland China as a first language and by Chinese people aboard as a second language. Applying the categories proposed in this theory perfectly to Mandarin Chinese may shed new light on Chinese grammar study and English language teaching in China. We hypothesise that the theory's three fundamental concepts—collocation, colligation, and semantic association—are applicable to Mandarin Chinese.

1.4 Theoretical framework of this thesis

The theoretical backbone of this thesis is Hoey's lexical priming theory for two reasons. First, we hypothesise that this theory can be expanded to languages other than English. As discussed above, lexical priming has grown from and is enlightened by the ideas of Sinclair, Winter, and Stubbs, amongst many others. Hoey, by accumulating and extending their ideas, proposed a fuller theory that accounts for several contentious issues, including such corpus linguistic phenomena as collocation, colligation, and semantic preference (association). Since it was first published in 2005, this theory has been mainly tested on English from many linguistic perspectives (as will be detailed in Chapter 2) and in a handful studies in other languages such as Turkish (Baker et al 2017) and Finnish (Jantunen 2017). As few studies have been conducted to account for the linguistic features of Mandarin Chinese, and as lexical priming theory has proven valid for English, we will now test how well it can be applied for Mandarin Chinese.

Second, the theory is directly applicable to L1/L2 learning and teaching. Michael Hoey has years of experience teaching English as a foreign language, and is therefore very concerned with the theory's applications for first and second language

acquisition. According to Hoey (2005), language is acquired through a range of speakers within the same community, the social context as a whole, and the quantity of exposure to the input environment; thus, what distinguishes language learners is not whether they are native or non-native speakers, but how their lexical primings come into existence. Native speakers encounter more possibilities to prime or encounter more primings, while L2 learners must establish their primings through fewer encounters. Language teaching and learning materials, as indicated in Section 1.2, do not often provide co-textual priming, i.e., priming made up by collocation, colligation, and/or semantic association. The lexical priming theory proposes that language learners should be exposed to authentic data wherever possible to reinforce existing or weaken inappropriate primings (here, we refer to primings that may sound unnatural to listeners who are native speakers).

Priming builds on insights gained from corpus linguistics, and applies the notion of psycholinguistics thereto. Psychology is invisible and works on a far subtler level than the re-occurrences of word patterns. Even Hoey admitted lexical priming can only be theoretically tested through laboratory experiment on human minds and neurons. However, by making an analogy between mental concordance and computer concordance, priming can be studied by examining corpus evidence to indicate "the kinds of data a language user might encounter in the course of being primed" (Hoey 2005:14). Corpus cannot unveil how the head neurons work in the process of individual primings, but indicates which primings are shared by large numbers of speakers in a language community. Indeed, psycholinguistics has shown corpus-based studies can provide results almost identical ($<95.0\%$) to results obtained from psycholinguistic experiments (Gries 2005).

This thesis is also guided by cross-linguistic studies between English and Mandarin Chinese (e.g., Wei and Li 2014, Xiao 2007; 2011, Xiao and McEnery 2006) that combine both quantitative and qualitative data collection and analysis methods. With respect to Xiao's work, the collocation and semantic prosody of English synonyms and their Chinese equivalences are examined. Xiao and McEnery (2006) restricted their studies to "result" group words by detecting data obtained from corpus. Subsequently, they analysed all concordance lines containing the "result" group words (e.g., *aftermath*, *outcome*, *consequence*) in English and their equivalencies in Chinese. Xiao's methodology—i.e., retrieving concordance lines containing research words from a large size corpus, and then specifically analysing lexical and syntactical behaviours of the extracted lines—provides general guidance for the present research project. Although Xiao contrasted English and Chinese from different angles, he did not offer a rational explanation for the selection of "corresponding equivalences." To fill this gap, the project will adopt a framework proposed by Altenberg (1999:254) to determine corresponding pairs across languages by building a parallel corpus and selecting equivalent patterns by calculating their mutual corresponding (MC) value.

Using Altenberg's framework, Wei and Li (2014) showed cross-linguistic equivalence should reside in corresponding patterns of words, rather than in single words as Xiao had subconsciously taken for granted. Cross-linguistic corpus studies by these researchers suggest English and Chinese should be contrasted at multiple layers, as looking at only one perspective does not provide a complete picture. Furthermore, previous studies have shown Chinese is much more complex than

English and may deserve more rigorous definitions of collocation and colligation.

1.5 The research questions

This section introduces the major research goals of this project, and the three research questions formulated to evidence its major assumptions.

1.5.1 Research goals

First, we will test how far the three fundamental concepts—collocation, colligation, and semantic association—can be applied to Mandarin Chinese, and whether doing so can advance or extend the theory. Second, we will identify the characteristic priming similarities and differences between English and Chinese, and how and to what degree these similarities and differences facilitate cross-linguistic study of the two languages. Third, we will examine the study's pedagogical implications and suggestions for L2 English teaching in the Chinese context.

1.5.2 The selection of research nodes

The present study will focus on an examination of a Chinese lexical item (*世界 shi4jie4*) and its English corresponding equivalent (*world*) to detect how Chinese may be primed in terms of collocation, colligation, and semantic association.

While the selection of the lexical item 世界 *shi4jie4* will be discussed in greater detail in Chapter 3, there were two main reasons. First, the lexical items *world* and 世界 *shi4jie4* occur with a similar frequency in the two languages, with *world* occurring in the British National Corpus over 500 times per million running words and 世界 *shi4jie4* occurring 560 times per million running words in the Chinese National Corpus. The frequent use of the node in Chinese implies numerous concordance lines will be found in various cases, which could offer ample data for its potential encounterable instances in Mandarin Chinese. In addition, the nodes' frequency of use demonstrates a strong comparability across the two languages, in terms of quantity. Second, the body senses of *world* and 世界 *shi4jie4* are shared by speakers of both languages, and are used to describe not only the material or profane sphere, but also the celestial, spiritual, transcendent, or sacred spheres. In the physical context, *world*/世界 *shi4jie4* represents the planet or universe in which we are living. In the philosophical context, classifying the concept of *world*/世界 *shi4jie4* has been a major concern of both Western (e.g., Plato, Hegel) and Ancient Chinese philosophy (e.g., Taoism). *World* in modern English evolved from the Old English *weorold*, a compound of *wer* "man" and *eld* "age" meaning roughly, the "Age of Man" (Wikipedia). However, 世界 *shi4jie4* in modern Chinese can be attributed to the translation of Buddhist scriptures, wherein 世 *shi4* in general refers to "time" and 界 *jie4* to "space" (Online Etymology Dictionary of Modern Chinese). In Buddhism, 世界 *shi4jie4* can thus refer to both the planet we are living on (in a narrow sense) and one's pre-existing, present, and future life (in a broad sense). The similar understanding of the external and spiritual world inherited by the two groups of speakers may demonstrate, to a large extent, similar senses embedded within *world*

and 世界 *shi4jie4.* Unlike words bearing strong, inherited, culture-dependent meanings, e.g., *dragon* and 龙 *long2*, due to their different interpretations and characteristic identifications, *world* and 世界 *shi4jie4* are likely to be culture-independent items whose semantic associations are unlikely to be presumably identified for culture influence.

1.5.3 Research questions

Based on the theoretical framework and major hypothesis that LPT could be applied to both English and Mandarin Chinese, we formulated the following research questions:

1. Can collocation, colligation and semantic association as defined in the Lexical Priming Theory account for the range of the lexical and syntactical behaviour of Chinese?

To answer this question, we will undertake a holistic phraseological examination and analysis of 世界 *shi4jie4,* regarding not only its collocates but also its colligation and semantic association, as defined in LP. The study takes 世界 *shi4jie4* as its starting point and focuses on frequently-occurring word sequences containing the keyword to provide a holistic picture at multi-layered contextual levels.

2. If the answer to the first research question is positive, in what ways do English and Chinese differ or show similarity with regard to the characteristic primings for lexical and syntactical behaviour?

This research question will be answered by undertaking a holistic examination and analysis of the English corresponding equivalence to 世界 *shi4jie4*, i.e., 'world'. We will start with the single lexical item *world* and focus on its phraseological corresponding equivalences to its counterpart; our ways of determining "equivalence" will be specified in Chapter 3. By doing so, we intend to find characteristic collocations, colligations and semantic associations of the two nodes.

3. Based on any lexical and syntactical similarities and differences that we found between English and Chinese, what advances or alterations could we provide for the Lexical Priming Theory?

Using LP as the starting point for our research, we seek to expand and advance the theory using evidence found by applying it to Mandarin Chinese. Through cross-linguistic comparison, we intend to advance the theory's potential to account for cross-linguistic similarities and differences between English and Chinese.

1.6 Definition of key terms applied in the thesis

1.6.1 Mandarin Chinese

Chinese, in the common sense, has many dialects. For historical or cultural reasons, it is spoken not only in mainland China, but also in many other geographical areas. In this study, Chinese refers to Mandarin (also called Putonghua or Standard Chinese). In addition to being the official language of mainland China, Mandarin is the major dialect family of Chinese, comprising pronunciations from the Beijing and Harbin

dialects, northern Mandarin grammar, and the vocabulary of modern vernacular literature. This thesis uses the terms Chinese and Mandarin interchangeably.

1.6.2 Definition of "words"

In most writing and printing conventions, an orthographic word is recognised as a string of letters placed between spaces. However, Mandarin has a distinctive writing system compared with Indo-European languages. A single word could be expressed with a combination of more than one character in Chinese (and vice versa); e.g., paint (a word) and 油漆 *you2qi1* (a combination of two characters and syllables)[2]. Thus, while this study uses 'word' as a starting point for analysis, it can be confusing to use "word" when describing Chinese; as such, we must clarify the notion of word for both English and Chinese at the preliminary stage. As discussed above, the project is concerned only with providing a holistic picture of one word pair—*world* and its corresponding Chinese equivalence[3] 世界 *shi4jie4*—so we have narrowed the definition of 'word' in this thesis to include only *world* and 世界 *shi4jie4*. Regarding clusters containing the two node words, Hoey's notion of "nesting"[4] will be used to help build Chinese characters into word combinations. With respect to collocates co-occurring with 世界 *shi4jie4*, we use "single lexical item" to refer a character (字 *zi4*) and "lexical item" to refer character combinations that could correspond to an English word found in a dictionary entry.

---

[2] The detailed typological features of Mandarin Chinese will be discussed in Chapter 3.
[3] The determination of "equivalence" and criteria for choosing the word pair will be detailed in Chapter 3.
[4] The notion of nesting will be introduced and discussed in the subsequent chapter.

1.6.3 The usage of "preference"/"avoidance"

Wording like "preference"/"avoidance," and "*** has a tendency/is likely to collocate (or colligate) with a particular other words" will be used in the data collection and interpretation chapters. Whenever this kind of terms applied, we refer to the psychological preferences of people using the word; referring to those preferences/avoidances in this way is a more economical way of describing the lexical behaviours found in English and Chinese. The primings indicated through the corpus data are not intended to suggest that words have preferences/avoidances themselves; rather it is a simplified way of expressing the idea demonstrated tendencies of the word (or word sequence) users.

1.7 Structure of this thesis

This thesis consists of nine chapters. Chapter 1 is an introduction, and explains the reasons informing this thesis, its theoretical framework, and the research questions it seeks to answer. Chapter 2 discusses the relation between lexis, grammar, and meaning since corpus linguistics has emerged, and then traces the historical development of the three major concepts used in the present study—collocation, colligation, and semantic association—and highlights how lexical priming has been applied to account for the existence thereof. Additionally, studies on the applications of and advances made using Lexical Priming Theory are reviewed. By way of critique on previous studies, the reserach gap the study intends to fill is identified.

Chapter 3 details the methodology used in the present study, and then describes the

corpora and analytical tools applied in the study. The research design is also specified.

The central chapters—Chapters 4 to 7 together report a series of results in various ways to show how corpus evidence can stimulate the hypotheses of the thesis. Chapter 4 reports the general collocational and colligational profile of English. In Chapter 5, we discuss the range of behaviour concerning collocations and semantic associations in different colligational situations. Collocations and semantic associations will be studied first at the clause level and then at the group level. How adequately the categorisation of collocation and colligation account for the behaviour of the node word will also be investigated.

Using data collected in Chapters 4 and 5 for comparison purposes, Chapters 6 and 7 focus on testing the validity and applicability of lexical priming for Chinese, taking our investigation one step further. We will first examine, in Chapter 6, the collocational and colligational behaviour of the Chinese node word, before investigating complex issues concerning the combination of collocation, semantic association, and colligation in Chapter 7.

Chapter 8 presents analysis and discussion, based on the evidence collected from Chapters 4 to 7. First, we discuss the feasibility of using LP to account for Chinese collocational behaviour. Second, we describe how Chinese is primed in terms of collocation, colligation, and semantic association; the unique features of Chinese primings are also discussed, advancing LP to a fuller and more satisfying theory. Last, we compare the characteristic priming features between *world* and *世界shi4jie4*

to identify implications for both L2 teaching and learning.

Finally, Chapter 9 presents this study's major findings and limitations, and identifies areas for future research.

CHAPTER2

Literature Review

2.1 Introduction

Since the principal aim of the present study is to examine the applicability of three fundamental concepts—collocation, colligation, and semantic association—utilized in lexical priming theory to both English and Chinese, their various characteristics need to be discussed and their historical development depicted. The Lexical Priming Theory (LPT) does not exist in isolation and was not "created" by any one scholar; rather, it is based upon previous studies and theoretical foundations laid by scholars in a range of fields. Therefore, this chapter will provide a brief chronological review of the contributions to our understanding of the concepts of collocation, colligation, and semantic association, with particular regard to the ways in which they have been defined, from when the ideas behind these concepts were first propounded until the final publication of LPT in 2005.

Section 2.1 outlines the major purpose of this chapter. Subsequent sections will detail where in this study the three concepts will be examined, as follows:

The base belief of this project is that grammar and lexis should not be separated or isolated from each other, but should instead be examined together. Thus, Section 2.2 will discuss the relationship between grammar, lexis, and meaning. It will also

discuss the development of corpus linguistics and how corpus linguists have taken advantage of computers and corpus analysis tools to process large amount of data to examine the grammatical role of English.

Section 2.3 introduces the ideas of and minds behind LPT, by referring to contributions made by previous researchers in defining the three concepts examined in the study.

Section 2.4 deals with the applications and advances provided by studies on lexical priming. These studies are generally divided into three groups: applications and advances gained from English; applications and advances gained from other languages (excluding Chinese); and applications and advances gained from Chinese. Relatively more studies have been undertaken on the use of lexical priming in English, while studies on Chinese are the least numerous of other languages, indicating a major gap that must be filled to "maturise" the theory. So, in Section 2.5, the literature gap depicted by reviewing previous work and the potential contributions of this project are discussed, and then summarise in Section 2.6, at the end of this chapter.

2.2 Grammar, lexis and meaning

LPT is a new theory developed by accumulating ideas proposed by corpus linguists on reversing the relationship between lexis and grammar. The position adopted in this section is in sharp contrast with the classic view on the relationship between lexis, grammar, and meaning.

2.2.1 Grammar, lexis and meaning

Bloomfield (1933) expressed the view that meaning was a "weak point" of language study. Two years later, Firth (1935) argued that studying a language without meaning is "sterile." His notion of collocation, as discussed early, serves as a connecting thread of meaning as function in context, in which there are syntagmatic relationships between words.

Tradition, according to Chomsky, holds that linguistics is a branch of cognitive psychology (competence) that can be based on intuitive data and isolated sentences. Early Chomskyan linguists thought that studying language in use (performance) is not interesting and that linguistics can be based on a variant of the Saussurian *langue* and *parole*. The main difference between the Chomskyan and neo-Firthian (Halliday, Sinclair) perspectives lies in their opposing views on how to treat naturally occurring data:

> The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of langue to account for masses of evidence that are hardly open to serious question. (Chomsky 1965: 19-20)

> Starved of adequate data, linguistics languished--indeed it became almost totally introverted. (Sinclair 1991:1)

Linguists in the tradition of Bloomfield or Chomsky seem to undervalue the connection between semantic and syntactic features. However, Sinclair (2004a: 6-7) suggested there is no distinction between meaning and form, saying "each meaning can be associated with a distinct formal patterning... There is no distinction between form and meaning... meaning affects the structure..." According to Sinclair, decisions

22

as to meaning are made initially in an abstract way; the final form is realised through lexical and grammatical ramifications, and default options that embody the rules of grammar. In this sense, grammar serves as the "management of text rather than the focus of the meaning creation" (Sinclair 1991: 8).

We must underline Chomsky's contribution to forming and legitimating the basic means of studying linguistics, even if the rationale on which we base this thesis is mainly neo-Firthian—or, to be more specific, Sinclarian. Indeed, Chomskyan and Sinclarian linguists all discover the truth; what makes them different is that they adopt different positions and describe things from different perspectives. The main difference may be that the former believe meaning can be analysed through conceptual analysis, while the latter believe meaning might best be determined by analysing text as a physical object.

2.2.2 Corpus linguistics and grammar

This corpus-driven study investigates corpus data as either freely available corpora or a self-built dataset. Thus, it is necessary to review the evolvement of corpus as a research methodology in linguistic study.

In traditional approaches to linguistics, language is understood as a set modular system consisting of phonology, morphology, syntax, and semantics. As corpus linguistics has become the mainstream and has developed from a "slab" to "bandwagon" (Svartvik 1996), it has shed new light on or even contradicted traditional views of the lexicon and grammar, especially the kind of grammar

initiated by Noam Chomsky's book *Syntactic Structures* (1957). Chomsky's view is that we should accept intuition as our primary form of evidence and reject linguistic performance as data; inevitably then, he rejected corpora as mere collections of linguistic performance in his early days. This view was adopted by following generations of theoretical linguists for many years, before corpus linguistics appeared in its present form.

Technology now allows us to discover patterns and combinations in raw textual data that are rather different from what we used to think, and we are advised to trust the text and let the language "speak out." Even in the early days of corpus linguistics, when computers had not yet been used for corpus analysis, some pioneers began to use authentic and natural texts to examine language from different perspectives and using different methods. For example, Alexander J. Ellis, author of *The Existing Phonology of English Dialects*, obtained data from over 800 persons, while the great early-twentieth-century Danish linguist Otto Jespersen filled a considerable number of drawers with authentic texts, and Charles Carpenter Fries obtained data from fifty hours of mechanically recorded conversations from a university community in the US. These examples seem to suggest that linguists began to pay more attention to the authenticity of language; their interest in studying real examples reflect their belief in the naturalness of language, which is insufficiently explained by either Chomsky's transformational-generative grammar or traditional grammar. In one of these studies, Quirk and Svartvik (1966) undertook an informant-based study to "devise a technique for establishing degrees and kinds of acceptability of different English sentences" (Svartvik 2007:17). They presented sentences to informants and asked them to carry out one of several operations thereon, such as turning the verb in the

sentence *They don't want some cake* into the past tense. The results showed that "24 out of 76 informants replaced *some* with *any*, and several others showed obvious discomfort over *some* with hesitations and deletions" (ibid). The informants' "discomfort" confirms the roles naturalness plays in a language. Grammatically, *They don't want some cake* is perfectly acceptable; however, the informants still felt discomforted and tended to believe there might be something wrong, something unnatural. As illustrated by the example in the Introduction chapter, which sowed the seed for this study, *learn the knowledge* is a grammatically correct but an unidiomatic combination, even if it bears a particular meaning when the "k" in *knowledge* is capitalised[5].

Another corpus study concerning the unsatisfactory nature of traditional grammar categories was Sinclair's (1991) investigation into *of*. Despite scattered distributions in different dictionaries, *of* seemed to be treated unanimously as a typical preposition. Prepositions are principally involved in combining with the nouns that follow them to produce prepositional phrases functioning as adjuncts in a clause, such as *We went to Liverpool **in the same week***. However, Sinclair found this principal usage of a preposition is not anything like the entry of *of* in the dictionary, which tends to combine it with receding nouns to produce elaborations of a nominal group—for example, *a small bottle of brandy*. Actually, the selection of *of* is far more complicated than traditional grammar suggests. The choice of *of* is determined by the choice of verb and is more sensitive to what precedes it than what follows it. Sinclair then concluded, with data evidenced by *of*, that a firm classification of uses and meanings should be made on the basis of ample evidence, and that grammatical

---

[5] "Learn the Knowledge" is an American website which assists taxi drivers update their knowledge of local address.

classification is sometimes not able to explain minor uses.

To address issues of fluency and accuracy in their construction of a grammar, Hunston and Francis (1998) examined whether traditional grammatical categories—object, complement, etc.—were adequate to explain the actual typical behaviours of verbs in the Collins COBUILD English Dictionary (1995), and found they were not. Hunston and Francis further explored the possibility of using another notion to analyse naturally-occurring texts, later defined as Pattern Grammar. Hunston and Francis found that some patterns are part of a traditional system of grammatical analysis, such as "V to-inf" and "V –ing" (*I started to follow him up the stairs* and *He stood up and began to move around the room*), and patterns of "V" and "V n," which is used to distinguish intransitive and transitive verbs in traditional grammar analysis; others patterns are not and may contain quite a number of elements, such as "V on v-ing" (e.g., *He insists on putting off the plan*), or be composed of elements represented by lexical items such as "V *way* prep" (*e.g., he talked his way into the post of chief costume designer*) (Hunston and Francis 1998:50). Even though they accepted that traditional functional categories can be treated as the primary means of coding verbs, Hunston and Francis suggested that these primary categories are too restrictive to account for all the patterns they found in their study, as the traditional categories could be used to adequately account for only a small number of verbs in English. They argued through example that "it is difficult to imagine, for example, how an analysis of *I'd love her to go into politics* as Verb + Object helps the learner to know how to use the verb *love* in this way more than the pattern description of V n *to*-ing" does. In addition to their findings regarding the patterns associated with verbs, Hunston and Francis also discovered

that sense and pattern are often associated with each other. For example, when *recover* has the sense of "to get better from an illness," it tends to have the pattern of V *from* n; when it has the sense of "to get back," it tends to have the pattern of V n (e.g., *Police recovered stolen goods*).

Corpus linguists have provided ample evidence that lexis and grammar are inseparable, and have attempted to "build together a grammar and lexis on an equal basis" (Sinclair 2004a: 164), aiming at the true integration of syntactic patterns and lexical patterns. A representative production made by this kind of attempt is Construction Grammar (Goldberg 1995; 2006, Ellis 2012), which posits that language could be described and conceptualised as constructions and entrenched as language knowledge in an individual's mind. In this sense, construction grammar combines cognition and language usage. Linguists supporting construction grammar also believe that "schematic constructions emerge from the conspiracy of memories of particular exemplars that language users have experienced" (Ellis *et al* 2010:26). This belief shows that, like LPT and other theories (e.g., pattern grammar and lexical grammar), construction grammar also talks of the inseparability of lexis, grammar, and semantics. What makes construction grammar and LPT more alike is that these two theories both try to explain, using insights from corpus linguistics and corpus data, how an individual's language knowledge is formulated through repeated exposure to contextualized instances. Memorized word combinations (collocation), certain structures (colligation), and reasons why these words are deemed appropriate (semantic association) all influence an individual's usage of language.

Although having a lot in common at the foundational level, construction grammar

and lexical priming are elaborated in slightly different directions. The former focuses on the establishment of prototypicality in a language; the latter attempts to account for the existence of notions used in corpus linguistics and text interaction phenomena (construction). The following section details the main focus and ideas proposed in LPT.

2.3 The Lexical Priming Theory

New ideas need incubation time, new techniques, new technologies, and an accumulation of previous ideas to finally make an impact. The LPT—as a backbone of the present study—did not occur by accident, but arose from earlier works. Since the focus of this study is to indicate how far LPT's three concepts—collocation, colligation, and semantic association (preference)—are applicable to Chinese, it is necessary to discuss why we find these concepts, as defined in LPT, to be valid.

2.3.1 Collocation

The development of the concept of collocation can generally be divided into three stages. Initially, collocation was used as a noun to describe a kind of linguistic phenomenon. The Merriam-Webster dictionary dates the origin of collocation back to 1605, when it meant "the act or result of placing or arranging together; specifically, a noticeable arrangement or conjoining of linguistic elements" (cited in Pace-Sigge 2013:12). Hoey (2005:3) pointed out that "collocation" was initially used by eighteenth-century amateur linguist Sir William Jones, while Doyle (2003)

commented that this notion was also cited in the 1928 edition of Webster's New International Dictionary. Both Webster's definitions and Sir William Jones's use of the notion suggest collocation has been slowly maturing since the 17th century.

The period from 1920 to 1957 can be seen as the term's second stage, in which collocation started to be used as an academic term for pedagogical purpose. Palmer (1933) established his list of frequently used English words while working in Japan as an English language teaching specialist. He carried out a corpus-based study on recurrent word sequences of English words, finally describing over 6,000 recurrent combinations that he published in his *Second Interim Report on English Collocations* in 1933. On the cover page of the report, Palmer wrote, "A collocation is a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts." Palmer then went on to construct what he would later call *Pattern Grammar* to show how constriction-patterns could be taught to facilitate English spoken/written production. Initially, Palmer termed his synthetic approach to the traditional parsing of sentences *mechanism grammar*; this was revisited by Hornby in his collaborative research and was used again and redefined by Hunston and Francis in 1996 as *pattern grammar*.

The third stage started in the 1950s and 1960s. With the onset of large, electrically-documented corpora and the development of analysis technology, collocation evolved slowly from a notion to a theoretical term. Firth (1957:194) was the first to point out that one "shall know a word by the company it keeps." Firth was not only concerned with the term collocation itself, but the relationship between meaning and collocation. Since Firth, collocation has had a number of definitions,

which can generally be grouped into visible and invisible ones. The visible definitions were statistical and textual ones, as described in various ways by Sinclair (1991), Kjellmer (1984), and Biber *et al* (1998), etc., who saw collocation as an observable phenomenon that could be reflected by concordances and computed by statistic scores, such as MI score and Z score. Other linguists, such as Hoey (2005), Partington (1998), and Halliday & Hasan (1976), believed collocation was invisible (psychological) but accepted that, to discover collocations, one must examine the statistical distribution of words occurring with greater than random probability.

The development of corpora stored in computers and corresponding analytic tools prompted new methods of assessing Kjellmer's (1998) distinctiveness; i.e., moving from intuitive to empirical means. One of the key pioneers of and contributors to investigating collocation was John Sinclair, who based his empirical research on corpora of naturally occurring language. One of his insights on the concept of collocation is that "words 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: 417-418). To determine whether language users have free choice in their use of words in a rule-set or grammatical sequences, or whether there is a tendency for people to choose prefabricated sequences, Sinclair (1991: 109-110) characterised two principles of interpretation—the open-choice principle (slot-and-filler) and the idiom principle. The open-choice principle involves the potential of generating sentences according to set (grammatical) rules, and can be seen as "an analytical process which goes on in principle all the time, but whose results are only intermittently called for" (1991: 114). The idiom principle, on the other hand, recognises the tendency for physical,

cognitive, or social phenomena to be associated, and for that association to be reflected in the language.

The idiom principle is manifested in the use of "a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair, 1991:110). In Sinclair's later work (1996, 2004a), he argued that this distinction is too sharp, even though it correlated the bilateral freedom of choosing a word and a grammatical structure. Sinclair further argued that some combinations are so tightly bound or lexicalised that they form a unit in which they behave as if a single item, e.g., *of course*. The *of* in *of course* is not the traditional preposition described in traditional grammar books and *course* is not a countable noun as categorised by dictionaries; its meaning is manifested in the phrase, *of course*.

Sinclair (2004a:31) and Stubbs (2001:27-28) further noted that different word forms of the same lemma have little overlap in their collocates. Sinclair noted there is very little overlap between the top twenty collocates of the plural *eyes* and the singular *eye*; *blue* and *brown* collocate only with *eyes*, while *caught* and *mind* collocate only with *eye*, as part of multi-word expressions to do with monitoring, visualising, and evaluating. Similarly, Stubbs (2001, 2009) found that different forms of *seek* co-occur with different collocates. Stubbs studied the top 20 frequent collocates co-occurring with different forms of *seek* in a 200-million-word corpus and discovered that *seek*, *seeking* and *sought* all shared the collocates *asylum*, *court*, *government*, *help*, *political*, and *support*; while the forms *seeks* and *seek* shared only one collocate: *professional*. In contrast, the word forms of *seeks*/*sought* and

seeks/*seeking* had no shared collocates. Stubbs further examined the contexts containing each form of *seek* and noted that *seek*, *seeking*, and *sought* often occur in texts on political and legal topics, with collocates meaning "help and support," but the word-form *seeks* is only distantly attached to this cluster. Likewise, Doyle (2003) showed that grammatically related forms of lemmata in scientific textbooks have few shared collocates; for example, *amplifier* and *amplifiers* have only three shared collocates, *circuit* and *circuits* have only two, and *shift* and *shifts* have none. Greaves (2007) and Warren (2007) argued that collocational analysis should be done on words rather than lemmata, on the grounds that each word has its own special collocational behaviour.

2.3.2 Collocation in LPT

The picture we have just sketched has in some respects shown that collocation is more than just two or more words occurring frequently in a context; rather, it is a key factor representing the naturalness of and pervasiveness in a language.

Hoey initially accepted what he, Sinclair and other linguists found for the term collocation, either statistically or textually. What attracted him later, however, was why collocation comes into being in the first place. This brought a "psychological dimension" (Pace-Sigge 2013:13) in play, changing collocation from a visible or observable phenomenon whose frequency and strength of co-occurrence could be computed with corpus analytical tools, to a psychological or subconscious process. Just as a corpus has a concordance, Hoey believed our brains also have a "mental concordance" (2015:14). He assumed priming might serve account for an

individual's collocational use, because our minds seem to sort out our language according to the statistical frequency (high or low) of collocates. This frequency is influenced by an individual's encounters with a word, the experience of which is either reinforced or weakened in the course of the individual's lifetime. Hoey termed this priming, and argued that it shapes the wording we use. The development of software and analysing tools has enabled researchers to identify recurrent patterning at the level of the structural framework and of individual collocations. LPT argues that lexis is complex and systematic, and that grammar is an outcome of lexical structure. Hoey started the theory by observing the pervasiveness of collocation at the word level, and then expanded his claims to grammatical functions (colligation, see Section 2.3.3).

As discussed above, Firth is frequently credited with having coined the term collocation in a modern linguistic sense; in doing so, he to some extent set the agenda for later studies. Researchers interested in collocation have picked up different aspects of Firth's ideas, thus coming up with different but related definitions of collocation, which Partington (1998: 15-16) has sorted into three groups: textual, psychological or associative, and statistical definitions.

Textual definitions tend to view collocation as "a consequence of the linearity of language" (ibid: 15). The representative definition for this group is Sinclair's: "Collocation is the occurrence of two or more words within a short space of each other in a text" (Sinclair 1991: 170)"

Hoey added that this definition does not reflect Sinclair's own use of this term and

can "result in a woolly confusion of single instances of co-occurrence with repeated patterns of co-occurrence" (Hoey 2005:3). What blurs the definition is that it does not specify how many words together are appropriate and what should be the proper length for a collocation. Sinclair himself did not use this definition in his later work; we introduce it here simply for the historical value of definition itself, rather than its actual application in reality.

Psychological or associative definitions could be used to explain "expectancies" (Firth 1957:195) that are realised through people's communicative competence. For example, in a given context, people may not know the meaning of a particular word; however, they may gain clues from the context surrounding the word and attempt to "expect" what the unfamiliar item might mean. Leech refers to this as "collocative meaning," which "consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment" (Leech 1974: 20). Through lifelong exposure to the meanings of words, native speakers form a general knowledge of word-meaning that becomes the basis for their expectancies of unknown words.

The third group, the statistical definition of collocation, is best represented by Hoey's (1991:6-7) comment that "collocation has long been the name given to the relationship a lexical item has with items that appear with greater than random probability in its (textual) context." Partington highlighted this definition, commenting that it is "a good working definition of the concept for those studying corpus linguistics, where large quantities of text can be made available for computer analysis" (1998: 16). In the era of favouring language study from quantitative

perspective and advocating the combination of linguistics and computer technologies, this definition seems suitable for investigating the "habitual associations" of a word with other items and obtaining lists of a word's frequent collocates. The statistical distribution of words is worth examining if one wants to discover collocation; however, as Hoey (2005: 3-4) claimed, "the definition says nothing interesting about the phenomenon, it gives no clues as to why collocation should exist in the first place" because it seems to suggest a methodology for studying collocation rather than describing the reasons underlying the existence of collocation.

Hoey then turned to the psychological/associative definition, such as that provided by Halliday and Hanson (1976:287) in their pioneering work on cohesion in English, wherein they refer to collocation as a property of the mental lexicon:

> Without being aware of it, each occurrence of a lexical item carries with its own textual history, a particular collocational environment that has been building up in the course of the creation of the text and that will provide the context within which the item will be incarnated on this particular occasion. (Halliday and Hasan 1976: 289).

Halliday and Hasan also highlighted the psychological underpinnings of collocation:

> The importance of collocation for a theory of the lexicon lies in the fact that at least some sentences (and this puts it cautiously) are made up of interlocking collocations such that they could be said to reproduce, albeit with important variations, stretches of earlier sentences. (2005:5)

However, they did not depict the relationship clearly, nor identify the cause-effect relationship between the psychological process and its outcome. Collocation is not characteristically built up "in the course of the creation of the text"; rather, it has previously been "mentally primed" (Hoey 2005: 8) for collocational use without the language user being aware. Hoey (2005:7) argued a good starting point for his theory was the pervasiveness of collocation, which was much noted by Sinclair (1991).

Hoey used two sentences— the first sentence from the main body of Bill Bryson's travel book *Neither Here Nor There* (1991) and Hoey's own contrived rewritten version thereof—to underpin the pervasiveness of collocation. The sentences are:

> *1) In winter Hammerfest is a thirty-hour ride by bus from Oslo, though why anyone would want to go there in winter is a question worth considering.*

> *2) Through winter, rides between Oslo and Hammerfest use thirty hours up in a bus, though why travellers would select to ride there then might be pondered.*

It is obvious that, though the second version maintains the meaning of the original and shares a number of words and lemma, most readers would consider the first to be natural and fluent, and the second artificial and clumsy. Hoey then claimed one of the obvious ways in which the two sentences differ is in the pervasiveness of collocations manifested in each sentence, as compared below using *Guardian* corpus data:

combination in Bill Bryson's sentence	times of occurrence	combination in Hoey's rewritten sentence	times of occurrence
in winter	507	through winter	7
by bus	116	in a bus	15
anyone would	122	travellers would	13
would want	573	would select	21

As the table clearly shows, Hoey's rewritten version of Bryson's sentence, though making use of existing collocations, has fewer occurrence times and does not interconnect. Hoey claimed the psychological explanation could account for the recurrent co-occurrence of words, and then suggested that the most appropriate psychological concept was priming, which has been discussed by researchers in the field of psycholinguistics. Neely (1976; 1977) connected the words 'priming' and

'lexical' in an experiment and found volunteers used less reaction time to recognise 'target' words if they were exposed to 'related' priming words; for example, the priming (related) word *body* provoked volunteers to recognise the target word *heart* more quickly. Prominent research by Scarborough et al. (1977:14) mirrored and reinforced Neely's experiment in many respects, in that it showed that a prior presentation of a "stimulus word" (priming word) could affect the recognition speed of the later presentation (target word). Both Neely and Scarborough et al. seemed to indicate that a single exposure to a related word (or non-word) would be stored in one's head and, once the related trigger was given, a word belonging to the same semantic field would be retrieved from memory. The focus of this psycholinguistic discussion was the relationship between the prime and the target, "rather than the word *per se*" (Hoey 2005:8). Hoey noted priming in his theory, calling it "a property of word and what is primed to occur is seen as shedding light upon the priming item rather than the other way round" (ibid). Halliday and Hasan referred to the process of collecting priming through encounters as forming a mental lexicon, but did not provide the underlying reason for why this phenomenon exists. Hoey (2005:9) complemented Halliday and Hasan and explained some characteristics of priming in his theory:

> Priming need[s] not be a permanent feature of the word or word sequence; in principle, indeed, it never is. Every time we use a word, and every time we encounter it anew, the experience either reinforces the priming by confirming an existing association between the word and its co-texts and contexts, or it weakens the priming, if the encounter introduces the word in an unfamiliar context or co-text or if we have chosen in our own use of it to override its current priming.

We quoted this at length to demonstrate that Hoey's use of priming is distinct from that of some psycholinguists (Neely, Scarborough *et al*). Priming, in Hoey's theory, stresses the word itself (*per se*). There is no priming word to trigger a target word;

priming exists at the moment the individual decides to use a word. Priming in Hoey's theory generally has three properties. First, priming need not be a permanent feature of the word or word sequence, and can change over the course of an individual's life. Second, since the priming of a word or word sequence can shift, it thus can also drift for "a number of members of a particular community at the same time, [providing] a mechanism for temporary or permanent language change" (Hoey 2005:9). This drift has been evidenced by Pace-Sigge (2013), who showed that the English spoken by Liverpudlians should not be seen as a dialect, as it only has a drift of priming shared by a particular community (Liverpool) that only differs from Standard English in the usage of some combinations. Third, primings nest and combine. Hoey used *in winter* to explicate this property, as collocations of this phrase are separate from its components (*in* and *winter*). *In winter* collocates with BE words; however, *winter* can also collocate with *the*, *during*, or *that*, and then form the combination of *the winter* and *during winter*. Another example indicated by Hoey is the word *word*, which collocates with *say* (*say a word*), which then collocates with *against* (*say a word against*) and then with *won't* (*won't say a word against*). Fourth, primings may crack. One of the causes of cracking is education, and is mostly realised through teaching-learning process. When teachers tell students that their initial words are incorrectly primed (e.g., *goed*) they display a potential crack in priming. Hoey claimed that such cracks "can be mended either by rejecting the original priming or by rejecting the attack on the priming" (Hoey 2005:11). However, not all cracks can be healed, which may result in uncertainty about priming and lead to linguistic insecurity. Finally, primings can be receptive and productive. Receptive primings, Hoey (2005:12) stated, "occur when a word or word sequence is encountered in contexts in which no probability, or even possibility, of our ever being an active

participant – party political broadcasts, interviews with film stars, eighteenth-century novels – or where the speaker or writer is someone we dislike or have not empathy with – drunken football supporters, racists, but also sometimes stern teachers and people of a different age group." Productive primings "occur when a word or word sequence is repeatedly encountered in discourses and genres in which we are ourselves expected (or aspire) to participate and when the speakers or writers are those whom we like or wish to emulate." We receive new or uncommon primings we may forget if we do not use them in our later life, or if they are cracked by factors such as education. The productive feature refers to the productive ability of individuals and is realised through the reinforcement and confirmation of primings.

To sum up, priming in the LPT deals with the priming item *per se* and can be transitory or (semi-) permanent subconscious mental concordance that relates words to each other. Priming, though existing within the individual first of all, might be changed by the contexts and environments to which we are exposed. However, Whitsitt (2005:298) stressed that the idea of priming "is very significant point. Hoey makes that our expectation, which may even explain why we have collocations, is not sustained by linguistic or semantic principles."

2.3.3 Colligation in LPT

Another important concept defined in lexical priming is that of colligation, an influential idea first proposed by Firth (1957:14):

> The statement of meaning at the grammatical level is in terms of word and
> sentence classes or of similar categories and of the inter-relation of those
> categories in colligation. Grammatical relations should not be regarded as
> relations between words as such – between 'watched' and 'him' in 'I watched

him' – but between a personal pronoun, first person singular nominative, the past tense of a transitive verb and the third person singular in the oblique or objective form.

Though Firth introduced the term colligation into modern linguistics, he did not distinguish colligation from grammar clearly. Colligation here is still conceptualised as a kind of post hoc grammar that describes or summarises situations observed from a language. Halliday, one of Firth's students and a famous linguist in his own right, used this notion in a rather different way in his book, *The Language of the Chinese 'Secret History of the Mongols'*:

> An analysis of certain features of the piece reveals that it is necessary for grammatical purposes to set up a unit larger than the piece (but smaller than the paragraph) …We must therefore admit the need for a unit larger than the piece, and the unit to be set up to meet this requirement will be called the "sentence." *Of the relation between the sentence and piece are that: 1, a piece ending in liau or jᵛe will normally be final sentence in the sentence; 2, a piece ending in sᵛi₂, ŋa, heu or san ŋgeu₂ will normally be non-final in a sentence; 3, a piece ending in lai or kiu may be either final or non-final in a sentence.*
>
> (Halliday 1959:38; the italicised section was cited by Langendoen (1968) as an example of Halliday's use of colligation)

Halliday's conception of colligation as a relationship between grammar and collocation was the foundation for subsequent research by Sinclair (1996, 2004a), Partington (2003) and Hoey (1997a, 1997b, 2005), who used the term in very similar ways. Sinclair illustrated the phenomenon using the word *lap,* which he found occurred only in prepositional phrases, such as "*her knitting lying in her lap, he lifted the cat off his lap*" (1992:14). He used colligation to explain the co-occurrence between the node word *lap* and the grammatical categories within which it is used. Sinclair's findings correspond to Firth's general rule that "each word when used in a new context is a new word" (Firth 1957:190). Sinclair consistently argued that "the normal carrier of meaning is the phrase" (2005; cited in Stubbs 2009:124) and

concluded that "generative grammars of the day would not actually generate the sentences that occur in open text, and the model was more of a metaphor for understanding the complexities of natural language as a whole rather than the blueprint for an operational grammar of a particular language" (Sinclair 2004b:114).

Sinclair's (1996) published work on colligation predated Hoey's by about one year; however, Hoey first articulated his conception of colligation in a 1994 lecture and an unpublished paper presented at the 23rd International Systemic-Functional Congress in Sydney in 1996. It does not matter who raised the idea first, since Sinclair and Hoey worked together for 14 years at the University of Birmingham and may have unconsciously influenced each other in one way or another; what matters is that their understandings of the concept are very similar, though Hoey's is slightly broader, incorporating matters of sequence as well as grammatical relations. Extending Halliday's formulation of the colligational relationship (quoted above) and believing colligation may go beyond traditional grammatical relationships, Hoey (2005: 43) suggested colligation could be defined as "the grammatical company a word or word sequence keeps, (or avoids keeping) either within its own group or at a higher rank." Hoey (ibid) detailed the following features of colligations:

1. the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank;
2. the grammatical functions preferred or avoided by the group in which the word or word sequence participates;
3. the place in a sequence that a word or word sequence prefers (or avoids).

From this formulation, colligational statements can be either negative or positive. Hoey (1997) noted, for example, that *consequence* rarely appears as the object of a clause, as in the following sentence:

> Unfortunately, it also had this tragic *consequence* that the baby became

grossly bloated.

According to Hoey, *consequence* appears in his corpus (*Guardian*) as the object of a clause in only 4% of all instances, whereas *preferences* and *uses* occur in this grammatical position in over one-third of instances. On the other hand, *consequence* occurs as a complement in 25% of all instances, much more often than is normal for other abstract nouns; *preference* and *use,* for example, serve this function less than 7% of clauses.

The infrequency with which *consequence* occurs as an object is an example of negative colligation, just as its abundant occurrence as a complement is one of positive colligation. This phenomenon, along with other words (such as *ponder*) and word sequences (such as *in winter*), was revisited by Hoey (2005: 44-47), who noted that *ponder* tends to avoid the passive voice, and that *in winter* tends to prefer a relational process whereas *in the winter* tends to prefer a material process.

What distinguishes Hoey's use of the term from that of other linguists is that he has suggested that words may be primed for a physical location (textual position, such as sentence initial position) and grammatical context, especially where the word is a polyseme. His observation led to the "drinking problem" hypotheses (Hoey 2005: 82):

> 1. Where it can be shown that a common sense of a polysemous word is primed to favour certain collocations, semantic associations and/or colligations, the rarer sense of that word will be primed to avoid those collocations, semantic associations and colligations of the rarer word but, proportionally, less frequent.
>
> 2. Where two senses of a word are approximately as common as each other, they will both avoid each other's collocations, semantic associations and/or colligations.

3. Where either (1) or (2) do not apply, the effect will be humour, ambiguity (momentary or permanent), or a new meaning combining the two senses.

He then examined the colligation of *consequence* when used in different senses (*consequence=result* and *consequence=importance*) to test his Hypothesis 1; discussion of Hoey's other two hypotheses falls outside the scope of this thesis. The contrast between the two polysemous uses of *consequence* with respect to a whole range of characteristic primings is displayed in Figure 2.1, below:

Figure 2.1 The contrasting characteristic primings of the two uses of *consequence*

	Cosequence (=result)	Consequence (=important)
Collocation with *any*	—	＋
Collocation with *of*	—	＋
Colligation with subject and complement	Positive	Negative
Semantic association with LOGIC	＋	—
Semantic association with NEGATIVE EVALUATION	＋	—
Pragmatic association with DENIAL	—	＋
Textual colligation with theme	positive	Negative

(Hoey 2005: 87)

"+" in this table means the uses corresponding to the first column occur in its primings, while the marker of "=" indicates that there are no uses occurring in the primings.

The figure shows that the two uses of *consequence* differ colligationally within subject and complement, with c*onsequence=result* having a positive colligation with the two grammatical functions and *consequence=importance* a negative colligation. Though it is obvious that Sinclair and Hoey drew on the work of Firth and Halliday to produce similar definitions, what distinguishes Hoey's definition is that he added a psychological dimension by saying a lexical item "may be primed to occur in or with a particular grammatical function. Alternatively, it may be primed to avoid

appearance in or co-occurrence with a particular grammatical function" (Hoey 2005: 43).

If lexical priming only operated with regard to these two notions, it would be of little or no theoretical importance. As concluded earlier, lexical choice and grammatical patterns are associated with the generalisation of meaning. A word or lexical item's meaning is realised through its association with other words.

2.3.4 Semantic association in LPT

There are always co-occurrences that cannot be accounted for by collocation only. Collocation is designated to account for word sequences that co-occur with high frequency, but has nothing to say about less-frequently-used word combinations. Indeed, the concept of semantic association is not Hoey's own, even though he named it and investigated its characteristic features; rather, it grew out of two other concepts (semantic prosody and semantic preference), which are discussed in the following section.

2.3.4.1 Semantic preference and semantic prosody

Both semantic preference and semantic prosody play an important role in Sinclair's structure of unit of meaning, with semantic preference describing features at the word level and semantic prosody describing speakers' motivation for speaking.

Sinclair and Hunston defined semantic preference as "the restriction of regular

co-occurrence to items which share a semantic feature" (quoted by Sinclair 2004a:142). The problem here is what is meant by "the restriction." It is argued in this thesis that semantic preference is not simply a restriction on the possible co-occurrences of an item, but an intuition or priming that pre-exists in people's heads, based on previously encountered co-occurrences. Stubbs (2002:49) described semantic preference as "a lexical set of frequently occurring collocates [sharing] some semantic features." Stubbs (2001) had earlier illustrated this by referring to *large*, which typically collocates with items from the same semantic set (e.g., *number(s), scale, part, quantities, amount(s)*) to form a semantic association of "quantities and sizes." Partington (1998) examined the semantic preferences of the item *sheer*, showing that it collocates with a number of items from specific semantic sets, including "magnitude" and "volume," "force" and "energy," and "persistence." Sinclair (2004a:32) found that, among the collocates of the node *naked eye*, almost all instances with a preposition at N-2 have a word or phrase to do with *visibility*, either at N-3 or nearby—e.g., *XX is invisible to the **naked eye***. Per his observation, these words and phrases meet the same semantic preference criterion of "visibility," which arises from the common semantic features of a range of collocates of the node *naked eye*. Semantic preference interacts with syntactic patterning (colligation), word class, and meaning. Partington (2004:145) noted that in the pattern <the sheer (noun phrase) of (noun phrase) > the semantic preference is for "magnitude/weight/volume" or "force/strength/energy," whereas in the pattern <preposition of means/manner (e.g., *through, out of, by, because of, by virtue of*) +sheer> the preference is for "persistence."

A word has meanings when associated with other words. For Sinclair, semantic

prosody was the reason a speaker chose a "lexical item" (2004a). The phenomenon

of semantic prosody was first described by Louw (1993) but was developed by

Sinclair in his later work (e.g., Sinclair 1987, 2004a). It has also been used by Stubbs

(1996, 2001) and Partington (1998, 2004), and has been criticised by, for example,

Whitsitt (2003) and Stewart (2009). Stubbs (1996) and Louw (1993, 2000) tended to

regard semantic prosody as two polarities: a positive polarity and negative polarity.

According to them, semantic prosody can be "a positive (favorable) prosody and a

negative (unfavorable) prosody, without finer-grained analyses of its specific

functions in context" (Wei and Li 2014: 108). A third type—neutral attitudinal

meaning—might be applied if needed. Sinclair's early study of *set in* (1991:73-74),

Stubbs's study (1995) of *cause,* and Louw's of *utterly* (1993:160), among many

others, seem to suggest a word or phrase carries a particular attitudinal meaning.

Stubbs's work is typical of and insightful about the study of prosody. The following

excerpt includes sample collocates to support Stubb's prosodic statement and

provides sentence illustration:

> 1. CAUSE: A *cause* is something that makes something happen. To *cause* something means to make it happen.
>> 1a Most frequently, >90%, the circumstances are *unpleasant*. Typically, what is caused is: an accident, cancer, concern, damage, death, disease, pain, a problem, problems, trouble.
>> 1b The circumstances can include a wide range of *unpleasant* things, mostly expressed as *abstract nouns*, such as: alarm, anger, anxiety, chaos, commotion, confusion, confusion, crisis, delay, difficultly, distress, embarrassment, errors, explosion, harm, loss, inconvenience, nuisance, suspicion, uneasiness.
>> 1c Frequently, the unpleasant collocates are *medical*: Aids, blood, cancer, death, deaths, disease, heart, illness, injury, pain, suffering, symptoms, stress, virus…
>> 1g…typical examples are:
>> the rush hour causes problems for London's transport
>> dryness can cause trouble if plants are neglected
>> considerable damage has been done to buildings
>> I didn't see anything to cause immediate concern
>> some clumsy movement might have caused the accident
>>> (Stubbs 1995: 247)

The prosody of the node word *cause* is claimed to be coloured by its environment. Whitsitt (2003) challenged this claim, arguing it is flawed both philosophically and in terms of its ability to cope with readily available counter-evidence, particularly where language functions metaphorically. He argued that "verbs like *alleviate, heal, relieve, sooth,* etc […] all habitually appear in the company of clearly unpleasant words, yet it seems clear that a word like *alleviate*, to take one example, certainly does not come to have an unpleasant meaning because of that company" (Whitsitt 2003:296-297). Hunston (2007:252) extracted examples of *cause* (see Figure 2.2) from the journal *New Scientist* to point out that "a word which is used in a certain way in most contexts is not necessarily used in that way in all contexts." Among the examples she examined, no particular prosody could be discerned regarding the entity brought into being. The causation identified by the researcher may help the research, but the things caused (e.g., instances of *1* and *4*) are in themselves neither desirable nor undesirable.

Figure 2.2 Extracted instances of *cause* in New Scientist journal (Hunston 2007:252)

1.	*They [reserachers] also searched for a second group of molecules called desensitisation proteins, which temporarily stop the receptor cells from responding to an odour.* **These proteins cause a smell to become less strong if we continue to sniff at it**. *The keys to the reserchers' success were antibodies which recognise…*
2	*When mounted a few micrometres above a chip, the probe and circuit form a capacitor.* **Any AC signal flowing beneath the probe causes a displacement current to flow through this capacitor**. *The value of this current changes depending on the amplitude and phase of the AC signal, enabling the signal to be imaged.*
3	*If signals are seen,* **how will we be certain that they are caused by dark matter particles**? *One way of finding out would be…*
4	*Whatever their views about the causes of the more mysterious amphibian declines, herpetologists agree about one thing:* **the chief practical problem is distinguishing genuine, long-term changes in population trends from fluctuations caused by short-term variations in weather**.
5	*The astronomers found that a part of the absorption spectrum called the Lyman alpha line… had been red-shifted by a factor of 30390 so that it appeared at visible wavelengths.* **They knew that this Lyman alpha line was caused by a galaxy** *because it matched the equivalent line in a faint, visible galaxy near to the quasar.*
6	*Irwin speculates that those stars are young – only a few billion years old – because* **they formed in a burst of star formation caused by a tidal interaction with the Milky Way**.
7	*This lost cold air is replaced by the warmer air in the room which must subsequently be colled when the door is closed again.* **The low specific heat and thermal conductivity of the filling will minimise the heat loss caused by convection while the door is open**. *…However, I suspect that the energy saved by filling an upright freezer with polystyrene foam is not significant.*

There are a number of disagreements surrounding the concept of semantic prosody; Partington might be considered representative of one branch, and Sinclair of another. While Partington and Louw tended to associate semantic prosody with a distinction between positive and negative or favourable and unfavourable meaning, Sinclair's considered semantic prosody an obligatory property of a unit of meaning. According to Sinclair (2004a: 175), "The semantic prosodies express attitudinal and pragmatic meaning; they are the junction of form and function." Sinclair (2004a: 175-176) offered effort as an example of this kind of semantic prosody. It is a countable noun in English, and so has a plural form (efforts). One of the frequent collocates of efforts is the word to, which follows effort as an infinitive maker; the corpus of The bank of English, in Birmingham, records 9,617 such occurrences. The data show the left-hand collocates of efforts to include a number of adjectives (e.g., hysterical, frantic, futile, and strenuous) and verbs and verb phrases (e.g., blunder, hamper, were overwhelmed, etc.) likely to indicate the failure of the efforts. However, by contrast and almost ironically, the right-hand contexts include a set of verbs that talk about creative action, such as please, revive, work together, protect, support, gain, raise, activate, kindle, help, give, save, etc. The left-hand surroundings tend to be negative in Louw or Partington's terms, whereas the right-hand surroundings tend to be positive.

Aside from his investigation of *efforts*, Sinclair has also conducted a study of the word *budge* (2004: 142-147), in which he pointed out that semantic prosody is a discourse function belonging to the word sequence as a whole—the unit of meaning—rather than a single word. The predominant collocate of *budge* is *refuse to*; although not grammatically negative, *refuse* in English can be reasonably considered

a lexicalisation of a kind of non-positive meaning. Other collocates are words ending in *n't* or *not*, which also create a non-positive environment. Sinclair then considered why people use *budge* rather than the more common verb *move*, noting that "something does not *budge* when it does not *move* despite attempts to move it. From the perspective of the person who wants something to be moved, this is frustrating and irritating, and these emotions may find expression, because this is the *semantic prosody* of the use of *budge*" (ibid: 144). This implies, but does not explicitly state, that Sinclair considered semantic prosody to be, to some extent, a psychological process. According to Stubbs (2009:118), in the 1960s Sinclair had asked how objective data and subjective meaning were related to each other. One of the key questions he asked was, "What is the relationship between the physical evidence of collocation and the psychological sensation of meaning?" His conception of semantic prosody appears to be a partial answer to his question. Sinclair (2004a) claimed that the semantic prosody in his unit of meaning serves as the root for people's collocational and colligational choice. It seems that his semantic prosody has both pragmatic and textual functions, thus blurring the boundary between semantics and pragmatics, even if sometimes there is no clear distinction between the two.

Realizing the limitations in earlier definitions proposed by both him and other linguists, Stubbs (2008:178) preferred to use the term of "discourse prosody" to distinguish semantics and pragmatics. He thus re-structured Sinclair's unit of meaning to explain the relations with which each term deals, as follows:

Table 2.1 Re-structured descriptions of Sinclair's unit of meaning (adapted from

Stubbs 2008:179, my italics)

1.	collocation	lexis	tokens	co-occurring word forms	*deals with words in the lexical field*
2.	colligation	syntax	classes	co-occurring grammatical classes	*deals with* relations between linguistic signs
3.	semantic preference	semantics	topics	lexical field, similarity of meaning	*deals with* relations between the signs and the world
4.	discourse prosody	pragmatics	motivation	communicative purpose	*deals with* the relations between the signs and the speakers

Hunston (2006) revisited the concept of prosody and claimed it is best to keep Sinclair's definition of prosody as a discourse function and as a definition that is not simply characterised as a choice between "positive" (favourable) and "negative" (unfavourable) or "good" and "bad." Holding that the use of language resides firstly in the individual user and taking the psychological ground into account, Hoey adopted an approach different from that of Louw and Partington and closer to Stubbs' definition. Instead of trying to split the less-direct and implied-meaning quality into sub-groups, Hoey (2005) used an umbrella term (semantic association) to bridge various forms of semantic prosody and discourse function.

2.3.4.2 Semantic association

Semantic association and Sinclair's term, semantic preference, may be seen as "interchangeable" (Hoey 2005: 24), but Hoey abandoned using Sinclair's term to eliminate occasional confusion between lexical preference and psychological preference. Because Sinclair used semantic preference to refer to the "meaningful outcome of the complex of collocational and other choices made across a stretch of language" (Hoey 2005:24), and because the central feature of Hoey's LPT is that of psychological preference, Hoey clarified his definition of semantic association:

> The definition of semantic association that we have arrived at is that it exists when a word or word sequence is associated in the mind of a language user with a semantic set or class, some members of which are also collocates for that user. (ibid)

Hoey's semantic association has four characteristics. First, it "is a necessary generalisation and appears to reflect a regular kind of lexical priming" (Hoey 2005: 17). In addition, there will be co-occurrences that cannot be accounted for in terms of collocations. In the case of *a three-hour car ride*, corpus data reveal that words representing NUMBER tend to precede *hour* and words representing VEHICLE tend to follow *hour*. This in turn forms a combination of NUMBER + TIME + JOURNEY + VEHICLE. It is not hard to find examples of *12-hour bus ride* and *five-hour coach ride*, but it is very rare to say *27-week flight* or to be more extreme, *150-light-year odyssey*. These combinations are rare and even odd; however, they could be created on the basis of the same semantic association. Maybe one day human beings will use light-year as a common time measurement. Hoey (2005:19) claimed this kind of collocation is not reflected in corpus because primings differ from person to person:

> …while there may indeed be semantic associations that on the basis of corpus evidence do not have corresponding collocations, these general semantic

associations (i.e., associations primed for many speakers of the language) may be based on local collocations (i.e., collocations primed for only a few speakers)   of   the kind that the average corpus is unlikely to detect.

The second characteristic of semantic association is that it combines and nests. Bastow (2003) noted that when *daylight* collocates with *broad*, it has a strong tendency of collocating with *in*. Then, the sequence of *in broad daylight* is likely to occur with "something bad happening, usually connected to crime or violence" (e.g., *… having been abducted and then stabbed in broad daylight…*).

The third characteristic is that semantic association is domain/register specific. For example, Bastow noted that *men and women* tend to collocate with *young* in the corpus of US defence speeches. The combination of *young men and women* in turn has a semantic association with COMPLIMENTS (e.g., *bright young men and women*); however, the collocational behaviour of the sequence *men and women* is rather different; one of its frequent collocates is *between*, which is not manifested in the US defence speeches corpus.

The fourth characteristic of semantic association depends upon certain grammatical conditions being met; more simply put, it is structures dominant. For location, it is the *to* PLACE and *from* PLACE structures; for measurement of time, it is the NUMBER *hours* (*away*) *from* structure (Hoey 2005: 33).

Hoey examined the collocates of *consequence* within a five-word span occurring in his *Guardian* corpus. One semantic association of *consequence* is LOGIC, which could be then divided into three sub-classes: the first referring to Necessity (unavoidable, inevitable, inexorable, inescapable, etc.), the second to the

DIRECTNESS or THE STAGES OF THE LOGICAL PRICESS (direct, ultimate, long-term, immediate, knock-on, etc.), and the third to the NATURALNESS or EXPECTEDNESS OF THE PROCESS (likely, predictable, possible, probable, natural, etc.). Here, we take the first sub-class as an example to explain the grammatically-tied characteristic of semantic association. One dominant collocate to the left of *consequence* is *inevitable*, which occurs 85 times in the corpus. However, this collocate does not appear to the left position of *consequence*; to be more specific, instances containing the combination of *inevitable consequence* is frequently used by Guardian writers, but *consequence is inevitable* rarely so. Similar instances can be fabricated with cognate forms of *inevitable*, such as

a. The ***consequence*** was inevitable;

b. Inevitably the ***consequence*** was that…

The adjective *inevitable* belongs to the sub-class (NECESSITY) of the LOGIC semantic association of consequence, so we assume that, if *inevitable* is grammatically restricted, words from the same semantic association should then have the same performance. Investigation shows that only three of 1817 cases occurring in the grammatical structure represented the "invented" example:

…a hazardous *consequence* is perhaps unavoidable
The second *consequence* is more relevant to the newspapers themselves.
The *consequence* was evident in the state of housing, schools, and hospitals.

The fabricated or invented examples of *a* and *b* may be praised and pronounced well by armchair linguists, but such structures are not commonly found in actual performance.

Hoey (2005) then extended semantic association to the discoursal level, arguing that

words might be primed in textual semantic association, which he defined as:

> Words (or nested combinations) may be primed to occur (or to avoid occurring) in specific types of semantic relation, e.g., contrast, time sequence, exemplification…

This claim might appear strange on first sight but is in fact simply an extension of the notion of semantic association. Just as a word or word sequence tends to co-occur with other words (collocation), a lexical combination may also occur as part of specific type of semantic relation, such as contrast, comparison, time sequence, cause-effect, exemplification, and problem-solution. In other words, textual semantic association extends the span of collocation.

Examining the textual contexts of 100 instances connoting *sixty*, Hoey discovered that *sixty* has a strong preference for occurring in contrast relations (42 instances) in newspapers, but avoids occurring in non-contrast comparison relations (16 instances). This finding implies that textual choices and local clausal choices are not always separate and are sometimes dependent on each other.

So far, we have detailed ways of how LPT define the three categories, namely collocation, colligation and semantic associations. In subsequent sections, we will showcase ways of how LPT is applied and advanced with a number of studies undertaken for English and other languages.

2.4 Applications and advances of LPT

It has been more than ten years since the official 2005 publication of LPT, since when various applications and expansions have been made to advance this theory in

many dimensions. This section will showcase those expansions and applications, based on the author's knowledge and access to materials. The relevant studies are grouped broadly into three categories: applications and expansions gained from English; applications and expansions gained from other languages (except Chinese); and applications and expansions gained from Chinese. With respect to applications, we refer to studies using LPT as the starting point for their investigations, such as the pedagogical usage for language teaching and learning. With respect to advances, we refer to expansions (or modifications) based on investigations of other languages, other genres, and claims proposed but not fully concerned within LPT.

2.4.1 Applications and expansions gained from English

This section examines some of the many studies undertaken on the English language, beginning with Pace-Sigge's usage of LPT for spoken language. While lexical priming has often been used to explain facts derived from written materials, little research has focused on spoken English. To fill this gap, Pace-Sigge (2013) tested the applicability of Hoey's theory to spoken English using spoken corpus evidence.

Pace-Sigge (2013:6) hypothesised that "speakers in a geographically restricted area, through their constant usage and exchange with local peers, have primings reflected in their speech." It might not only be their grammatical or lexical differences that define a set of speakers as a separate speech group, but also their different ways of using the same lexicon. Pace-Sigge's research was based on a thoroughly-known corpus by the author—SCO (spoken Liverpool English or, in Pigger's term,

Scouse)—which was in turn based on casual conversations collected by the author from 2002 to 2006. It contains conversations involving over 50 Liverpudlians of both genders from ages eight to 80. Another three corpora were used for comparison purpose. The first was the British National Corpus Conversation Subcorpus (BNC/C), which contains some four million words and is "highly comparable to the SCO corpus in age and gender distinction" (Pace-Sigge 2013:7). The second corpus was the Collins Bank of English (BoE), a UK-spoken sub-corpus containing 9.2 million words. The final was the London Borough of Hackney (LDN-H), containing over 205,000 words. The gender balance of the third corpus is exactly 50:50 and its informants' general demographic information (age, social background[6],etc.) is similar to that of SCO.

Results from the SCO were then compared with transcripts collected throughout the United Kingdom. What Pace-Sigge tried to show was that lexical priming can be used to explain frequently-occurring "formulaic chunks" in spoken language. He hypothesised that "in casual spoken Liverpool English, it is not just the traditional criteria that identify a particular variety of language as a dialect… one variety of English may also differ from the recognised varieties of English in respect of systematic variations" (2013:4). These variations are realised through the use of distinctive collocations, colligations, and semantic associations, which are major terms examined in LPT.

Of Pace-Sigge's comprehensive studies on clusters used in Scouse, we only here

---

[6] Here refers to the socio-economic background, which is classified according to *National Statistics Socio-economic Classification*. This criterion is also used by the UK Office of National Statistics. Informants of children or young adults are not classified in both SCO and LDN-H corpora.

illustrate the comparison between *yeah* and *yes*. The first comparator may be disregarded and normally treated as a form of *yes* by non-native speakers. However, a detailed study of *yeah* and *yes* shows a wide range of uses. First, *yeah* occurs significantly more often than *yes* in both SCO and BNC/C, with four times and over twice as many occurrences in the two corpora, respectively, as *yes*. Furthermore, *yes* and *yeah* form different clusters, as presented in Figure 2.3, below:

Figure 2.3 Top clusters of *yes* and *yeah* in the corpora

SCO cluster YEAH	tot.	%	SCO cluster YES	tot.	%
YEAH – YEAH – YEAH	41	2.6	YES OF COURSE	3	3.0
OH YEAH – YEAH	20	1.3	YES YES FOR	2	2.0
THAT'S RIGHT YEAH	17	1.1	YES YOU CAN	2	2.0
YEAH THAT'S RIGHT	10	0.6			
BNC/C cluster YEAH	**tot.**	**%**	**BNC/C cluster YES**	**tot.**	**%**
YEAH YEAH YEAH	1015	1.7	YES, YES , YES	453	2.5
YEAH I KNOW	990	1.7	OH YES YES	281	1.6
YEAH YEAH I	601	1.0	YES I KNOW	238	1.3
YEAH . I MEAN	544	0.9	YES , THAT 'S RIGHT	232	1.3
YEAH BUT I	540	0.9	YES YES I	223	1.2
BNC/W cluster YEAH	**tot.**	**%**	**BNC/W cluster YES**	**tot.**	**%**
YEAH I KNOW	18	1.3	YES OF COURSE	423	2.2
YEAH HE SAID	11	0.8	YES I KNOW	245	1.3
YEAH SHE SAID	11	0.8	YES I DO	189	1.0
BoE cluster YEAH	**tot.**	**%**	**BoE cluster YES**	**tot.**	**%**
YEAH. YEAH. YEAH	19,022	21.8	YES. YES. YES	14,118	12.2
YEAH. YEAH. AND	3,643	4.2	OH YES. YES	4,002	3.5
YEAH. YEAH. SO	3,162	3.6	YES. YES. AND	3,793	3.3
YEAH. I MEAN	3,117	3.6	YES THAT'S RIGHT	3568	3.1
YEAH. YEAH. I	2,953	3.4	THAT'S RIGHT. YES	2,277	2.0
LDN-H cluster YEAH	**tot.**	**%**	**LDN-H cluster YES**	**tot.**	**%**
YEAH YEAH YEAH	111	4.2	SHE DOES YES	5	4.0
OH YEAH YEAH	49	1.9	I SAID YES	5	4.0
YEAH IT WAS	37	1.4			

(Pace-Sigge 2013: 73)

Though some overlaps appear, *yeah* and *yes* generally form different sets of clusters on the whole, which Hoey (2005:59) would describe as *nesting*:

> The property of nesting of primings is an important one in that it allows us to go some way beyond certain kinds of grammatical description. In particular, it helps us explain the existence of grammatical structures in apparent free variation.

The findings listed above, though only a small part of Pace-Sigge's comprehensive study, confirm his hypothesis that "all words and sets of words that diverge from the average should be specific to the speech community of Liverpool speakers" (ibid: 4). While clusters containing *yeah* and *yes* are frequently recorded in both SCO and the comparator corpora, their particular frequencies of use, characteristic semantic associations, and nestings of key items diverge. The patterns used by Liverpudlians are not idiolects, greatly different from other parts of the UK. However, the evidence for lexical priming in SCO shows different degrees of likelihood of the same pattern being used; their prominent use by Liverpool speakers can be seen as evidence of a speech community's patterns having becoming "self-re-enforced" (Pace-Sigge 2013: 179), which is consistent with the basic claims in LPT that members of a community can to a degree influence (prime) each other, and that those primings will in turn be mutually reinforced by the members of that community.

While Pace-Sigge attempted to apply and expand LPT using data derived from spoken English, K.J. Patterson[7] (2016) adopted it to account for English metaphoric features. Her paper addressed issues concerning current systems of categorisation and measurement of linguistic metaphoricity. Adopting two alternative theories—Theory of Norms and Exploitations (Hanks 2004) and LPT (Hoey 2005)—Patterson proposed LPT is valid and applicable for the explanation of linguistic norms and exploitations involved in metaphor: "The introduction of an extended theory [LPT] involving our psychological associations with words could possibly offer an explanation for how we recognise norms and exploitations in the first place" (2016: 245).

---

[7]This paper was published online on December 5th, 2014 and then was collected in the Journal of Psycholinguistic Research in 2016. This is cited as Patterson (2016).

Patterson summed up two major problems arising from the assignments of meaning to metaphor found in previous research in this field. One is the problems of Metaphoric and Literal Meaning. The distinction between these two is not always clear-cut, and cognitive linguists may focus on the "individuals' interpretation of truth or literality," while lexicographers may "place priority upon the word in context, and the surrounding language which is capable of defining that word's literality" (Patterson 2016: 242). A typical sentence using the metaphor *a bleeding heart* ("*My heart bleeds for them, literally bleeds.*") explains the inadequacy of explaining metaphor in terms of the aforementioned notions. As Patterson argued, "Without the word *literally*, the dependant clause stands along would probably unquestioningly be labeled as metaphorical." However, the word *literally* claims a truth for the wording. Knowing this, we still interpret this sentence as a metaphor, because we clearly understand that a person whose heart were, in fact, bleeding would not be able to speak.

The second major problem addressed by Patterson is that of Secondary Meaning, which relates to the "abstract level of semantic relationships" (ibid: 243) and is the deeper meaning that may be denoted by metaphor. Patterson considered Hoey's "drinking problem" and argued, quoting works by Philip (2011) and Hanks (2004), that the second meaning could be explored based on a mergence of collocation, colligation, semantic, and pragmatic associations. The figurative sense (or second meaning) can be activated through collocation and colligation at the word level. The metaphoric meaning could also be predicated by word knowledge, presumably formed in the readers' mind through cumulative encounters.

In accordance with the hypotheses of lexical priming, Patterson examined the non-metaphoric and metaphoric uses of the keyword *kindle*, in terms of its colligations, collocations, semantic, and brief pragmatic associations. Reading through each concordance line containing the node, Patterson found that the non-metaphoric usage of *kindle* (as a transitive verb) co-occurred with words such as *fire*/*flame*/*torch* as a direct object, while the metaphoric usage of *kindle* co-occurred most often (75% of occurrences) with abstract objects that could associated with a meaning of *human emotion* (*love*, *desire*, *thoughts*, etc.). Such typical colligational structures as (*into*) *flame*/*fire of* + *abstract noun* were also detected. The results in Patterson's study show that collocates co-occurring with a metaphor may be also grouped semantically and colligationally. The interpretation of a metaphoric meaning could be aided though higher frequency and conventionality. The motivation behind the uses of metaphor could be explained as a kind of priming. The metaphor might be learnt and understood through encounters with it in different contexts, which in turn form part of our knowledge of this metaphor co-occurring with other words. In other words, metaphors are primed in the same way as other features of language.

The third study discussed here introduces a practical use of LPT. Jeaco's (2017) design for a concordancer (the Prime Machine), created for his doctoral degree project, is firmly based on categorisations noted in LPT. Appreciating some of the struggles faced by English teachers and learners in China, Jeaco designed the machine to help learners and teachers explore various features indicated in LPT, without having to know the theory as explicitly as researchers do. Jeaco emphasised that LPT brings together a range of linguistic features that, while important, are difficult for language learners and teachers to understand and master, especially

collocation, textual colligation, and semantic association. Based on concepts defined for the linguistic features in LPT, Jeaco designed his learner-friendly software to help individuals explore differences between words and phrases, notice textual colligation, co-text and contexts, and identify the features of the patterning of words and phrases. In doing so, the Prime Machine aims to help the leaner/user acquire a comprehensive knowledge of a lexical item, including the collocates frequently co-occurring with it, the sentence position it is likely to posit, and the grammatical structure within which it tends to occur. The Prime Machine brings together such linguistic features as lexical and grammatical patterning and textual and semantic patterning (like LPT), and then shows them on one screen to facilitate language teaching and learning.

Other studies have also made use of LPT for English. For example, Pace-Sigge (2017) considered priming when describing turn-taking in casual talk, while Duguid and Partington (2017) looked at what they termed "forced priming" in transdiscoursive political messaging, etc. Compared with the plentiful research carried out on English, fewer studies have examined the applicability of LPT for other languages. Among those few, we illustrate here studies that have made a good attempt at expanding LPT, not only by testing its validity, but also by extending it to claims not concerned fully with the theory.

2.4.2 Applications and expansions gained from other languages (except Chinese)

The first study reviewed here was by Baker *et al* (2017), in accordance with Hoey's claim of "a drift in priming" (2005:9). Hoey proposed that priming does not need to

be permanent, and could change over time as a language changes. Thus, time is clearly an important factor influencing the acquisition of primings. However, Hoey did not pay enough attention to tracing *drifts* in priming over time. To advance the theory, Baker *et al* (2017) filled this gap by looking at how Turks and Ottomans were referred to in seventeenth-century British English texts; as language changes rather slowly over time, a larger-sized corpus covering a long period was used in their study. Based on changes in English representation of Turks and Ottomans across the century, Baker *et al* concluded that primings are capable of drifting over time, yet the pattern of diachronic change is rather complex and "may be the result of new primings being added… But drift may also be expressed by receptive primings becoming productive, and vice versa" (2017:61). Another important feature of priming Baker *et al* noted is that major social pressures, such as religious anxiety and reactions to foreign policy, could influence primings to various extents.

While Baker *et al* focused on the diachronic change of priming over time by examining the evolvement of English representations of Turks and Ottomans, Jantunen (2017) examined the morphological priming of time expression in a case study of the Finnish word, *kello* (watch, time, o'clock). The study was motivated by the difficulties facing language learners whose target language is rich in morphology and morphosyntax (e.g., Finnish, German). One influential factor is the inflectional system between source and target languages. Few previous studies of learner phraseology have "taken into account the role of morphology in learning phraseology" (Jantunen 2017: 253) and most of those have dealt with English, which is not as morphologically rich as Finnish and similar languages. To fill in this gap, Jantunen undertook a holistic analysis of phraseological units containing the

keyword *kello*. The theoretical framework, drawn from LPT, was the paradigmatic morphological priming (PMP), which Hoey (2004: 24) discussed when he defined grammatical priming as "the grammatical category a word belong to" and continued that "What we account as grammar is the accumulation and interweaving of the primings of the most common sounds, syllables and words of the language" (2005:159). PMP argues that not all word forms are equally stored in the mind; some are core items in a paradigm while others are peripheral or even non-existing (Karlsson 1985, 1986). Based on Karlsson's argument, Jantunen assumed that certain core inflectional forms and their associations are also stored in the mental lexicon. The result reveals that learners of Finnish tend to choose the morphologically easiest of available grammatically acceptable alternatives. The study also shows that while learners' primed the target language "lexically deviantly, they primed it differently morphologically" (Jantunen 2017: 268). Furthermore, the author also revealed, concentrating solely on one lexical item, that even a single phraseological unit demands several contextual skills, including collocation and morphological associations and semantic association.

The two above studies provide new insights on LPT by examining languages that are distinctive from English. First, they show the applicability of some claims proposed in the theory for other languages, which moves LPT a step closer to universal usage. Second, the theoretical validity verified in the two studies by examining two different languages offers additional insights into language teaching and has significant pedagogical implications.

2.4.3 Applications and expansions gained from Chinese

LPT has not been as fully explored and tested with Chinese as it has with English, with Hoey and Shao (2015) being among the few researchers to address the issue. By looking at several examples of both English and Chinese, Hoey and Shao (2015) showed that English and Chinese share properties used in lexical priming, and that the two genetically different languages use a shared set of concepts and procedures. A general description of their study and its related results are outlined below.

First, they tested whether Chinese showed evidence of semantic association. By examining the data for 好 *hao3* (good) occurring with 法 *fa3* (way, method), Hoey and Shao found the two had a semantic association with DIET, as reflected by collocates such as 减肥 *jian3fei2* (lose weight) and 瘦身 *shou4shen1* (go on a diet), as well as with EDUCATION, as reflected by words such as 导入 *dao3ru4* (lead-in). This could suggest that 好 *hao3* and 法 *fa3* form a nested combination; if so, this is another property shared by the two languages.

They then looked at the polarity of clauses containing 后悔 *hou4hui3* (regret) + 事 *shi4* (a general word for matter, thing, and affair) and found 37 occurrences in positive clauses and 38 occurrences in negative instances. This is strikingly different from the 9:1 ratio between positive and negative in English clauses found by Halliday and James (1993). Finally, they tested the applicability of pragmatic association in Chinese by re-examining the combination *hou4hui3* + *shi4* and found that 31.6% of negative instances concerned a positive pragmatic association with

'making a suggestion/giving advice'. Though their work was very preliminary, it suggests the four concepts used in the LPT in English—collocation, colligation, semantic association, and pragmatic association—are also applicable to Chinese.

In addition to Hoey and Shao (2015), Ge (2014) also studied the use of LPT in Chinese by exploring the relationship between translation and language change in a Chinese-English translation context. A growing body of research acknowledges that translation influences language change; however, how translation affects target language change and whether the target language is influenced continuously over time remain to be answered. According to Ge's review, there is no theory that can be used to explore Chinese language change due to translation activities.

Ge's assumption is based on an expansion made by Hoey himself in 2011. By inviting three native Portuguese students to translate the first sentence from Bill Bryson's travel book, and then comparing the versions with a multilingual corpus including both English and Portuguese, Hoey found that the apparent exact equivalents of *in winter* and *no inverno* (*in winter* in Portuguese) were not primed exactly the same. For example, *in winter* was normally found at the beginning of a sentence, while *no inverno* showed little such tendency in Portuguese. In the translation process, translators were primed with the textual colligation of the target language by using *no inverno* at the initial position of a sentence. Though a larger corpus is needed to provide ample data, based on the data examined to date, Hoey proposed the language shift from source to target language might be a significant indicator of the role of translation in language change.

Reviewing the literature on the Europeanization of Chinese, Ge proposed that this language change could be accounted for using LPT. She further proposed that several modifications need to be made when taking translation as an approach to studying language change. First, a historical English-Chinese parallel translational corpus should be constructed for the analysis of lexical and grammatical change, to discover the possible influence of translation on target language change. Second, a comparable corpus should be included in studies to verify the hypothesis. Although Ge provided constructive insights by reviewing Hoey's assumptions regarding using LPT to account for language change influenced by translation, she did not undertake a practical corpus study to support her hypothesis.

Sections 2.4.1-2.4.3 discuss some of the research using LPT as a theoretical backbone or investigating claims proposed but not fully addressed in the theory. The review shows that LPT has been explored with plentiful work on English; however, relatively few studies have been undertaken for other languages, save for the notable exceptions discussed above. Of the few researchers to focus on Chinese, only Hoey and Shao (2015) and Shao (2017) have undertaken corpus studies of both conceptual and abstract associations defined in LPT. Others—including Ge (2014), Su (2012) and Li (2006)—have mainly addressed issues at the theoretical level by either proposing assumptions for explaining Chinese linguistic features using LPT or reviewing basic ideas and concepts defined in LPT. Examining Chinese through a combination of both corpus data and theoretical assumption is a necessity.

2.5 Applications and expansions of LPT in a new dimension

As shown above, LPT has been applied to account for a number of linguistic features. Expansions and advances have been provided either by applying the theory to a new linguistic field (e.g., spoken English) or offering ample data for claims proposed but not fully discussed in the theory (e.g., morphological priming). A review of the literature applying LPT to languages other than English (e.g., Finnish) implies that examining a language genetically different from English could generate new insights.

This thesis explores the concepts and categories noted in LPT to provide more evidence of LPT's ability to explain the linguistic features of Chinese and, based on that evidence, to determine whether LPT can be further advanced or expanded to address the characteristics of Chinese.

The issue of applying LPT to Chinese was addressed by Hoey and Shao (2015) through a preliminary study of several notions defined in LPT—collocation, semantic association, and pragmatic association. Their findings, based on corpus-based studies, convincingly show the possibility that the characteristic features of Chinese could be demonstrated with corpus and accounted for with psycholinguistic theory, such as LPT. However, Hoey and Shao (2015) were not able to extend their study to either colligation or the integration between colligation and other conceptual notions. They restricted their study to observing a given word or a cluster at a particular level of description (collocation, semantic and pragmatic association), rather than providing a holistic analysis of lexical items appearing in more complex patterns, such as in interactions between a set of collocates and colligations at two or more levels. In this project, we aim to fill this gap by

examining the given lexical items at three contextual levels—collocation, colligation, and semantic associations. This holistic study is undertaken by examining collocations and semantic associations across three colligations: the grammatical function within which the node serves; the position in which it is placed in a nominal group; and the position it occupies in a sentence.

Another goal is to compare the priming features across languages (English and Chinese). Corpus construction and the use of corpora for learning and teaching English in China have been greatly developed in recent decades. However, little work has been done on contrasting and comparing the two languages from a theoretical perspective. Corpus researchers in China seem to overemphasise using corpus as a methodology; such studies, though published in various journals, are "often accused by scholars in other linguistic disciplines of being strong in data but weak in theorising" (Wei 2015: xiii). Methods and theories can hardly be divorced or discussed separately as they are "mutually informed" (ibid: xiv). There are, however, a few researchers who have combined methodological issues relating to corpus use with attaining more evidence for a specific theory. Among these, the works of linguists Richard Xiao and Naixing Wei have provided great insights for the theoretical framework and methodology for this study.

Xiao has contributed a lot to the contrastive study of English and Chinese and has provided constructive suggestions for translation theories. For example, Xiao and McEnery (2006) examined three groups of near synonyms in English and Chinese—the CONSEQUENCE group, the CAUSE group, and the PRICE/COST group—to investigate patterns related to the collocational behaviour and semantic

prosodies of the words in these groups in Chinese. These features were then contrasted with English. They found four near synonyms in the CONSEQUENCE group in English (*consequence*, *outcome*, *result* and *aftermath*), the equivalent Chinese synonyms for which were 结果 *jie2guo3*=后果 *hou4guo3* (*result/outcome* = *consequence/aftermath*). Other alternatives were then added, including 成果 *cheng2guo3* (achievement) and 硕果 *shuo4guo3* (great achievement), which tend to express positive meanings, and 苦果 *ku3guo3* (a bitter pill to swallow) and 恶果 *e4guo3* (evil consequence), which tend to express negative meanings. Xiao and McEnery then arranged the words on a semantic continuum running from positive to negative—first the English synonyms (i.e., *outcome/result, consequence,* and *aftermath*) and then their Chinese equivalents (*shuo4guo3, cheng2guo3, jie2guo3, hou4guo3, ku3guo3/e4guo3*) (Xiao and McEnery 2006: 111-113). Compared with English, they noted, "Chinese is more sharply divided between the clearly negative and positive ends of the continuum" (ibid: 114), because Chinese has less morphology (e.g., *consequence* vs. *consequences*) that could attract different collocates, thereby affecting their semantic prosody slightly. Some of the significant contributions of their research are that Chinese has collocates and semantic prosody and that collocates of Chinese can be divided into different "meaning categories," a term very similar in meaning to LPT's semantic association (preference). Furthermore, they showed that translation equivalents (e.g., *consequence* vs. *hou4guo3*) display similar collocational behaviour and semantic prosodies and have the same "meaning categories" indicating ACTIONS associated with the affected target, such as *accept*, *avoid* etc..

Xiao and McEnery's work inspired this project by demonstrating the existence of

collocation and semantic prosody for Chinese; however, after encountering different theories and further reflections on Chinese, we find two shortcomings. The first is that they neither fully addressed the characteristic features of Chinese words and the differences/similarities between the two languages in advance nor discussed the definition of "word" across the two languages. With respect to the structural complexity of words, Mandarin is relative simpler than English. A typical word in Mandarin is not made up of morphemes, but is, rather, a single morpheme—for example, 书 *shu1* (book). Some "words" may have a strong tendency to be segmented while others may not. *Shuo4guo3*, for example, is a word with strong splitable characters, with 硕 *shuo4* (large) having a strong tendency to be collocated with 大 *da4* (big) to mean 硕大 *shuo4da4* (huge). Similarly, *guo3* 果 is a general term for fruit that can be collocated with *ping2* to mean 苹果 *ping2guo3* (apple), or with *huo3long2* to mean 火龙果 *huo3long2guo3* (pitaya). *Shuo4guo3* has an inherent positive meaning, so it is not surprising that *shuo3guo3* occurs in a context associated with positive meanings in Xiao and McEnery (2006).

Another issue with Xiao's research and other contrastive studies related to methodology, in that they decide equivalence either based on the researchers' own experience of a language or with reference to information from bilingual dictionaries (Wei and Li). This methodology gives rise to two questions. The first is, what is the criterion or measurement used to determine whether there is equivalence between words in different languages? The second is, who is authoritative enough to decide the degree of equivalence? Consulting a bilingual dictionary may be helpful but cannot be treated as a sufficient criterion for assuming equivalence, since words are primed with different meanings according to the words they collocate with, their

occurrence in different positions in a clause, and the different domains in which they appear. These variables can never be reflected adequately in a dictionary. Similarly, it is also not appropriate for any one individual to determine equivalence based on their experience and knowledge, as neither is reliable. Moreover, individuals' judgment is inevitably affected by their education and amount of exposure to the source and target languages, meaning their decisions on language are made subjectively. Wei and Li (2014) were aware of and dealt with this issue by adopting Altenberg's (1999:254) [8] Mutual Correspondence (henceforth MC) value. As a reference criterion, MC value provides an indication of the translatability of words across two different languages by computing the degrees of correspondence held by the words. According to Wei and Li (2014: 107), MC values serve as an "ad hoc reference criteria with which we select translation pairs for investigation." While the indicator may not determine an exact equivalent between two languages, it provides a reference for selecting frequently recurring corresponding pairs. The computation of this value requires establishment of a bi-directional parallel corpus (or database in the present study, because the data analysed are concordance lines extracted from general corpora, which are rather smaller than typical corpora, both in size and types of genres). The procedures for establishing this kind of dataset will be detailed in the next chapter.

So far, we have discussed the major issues we shall attempt to address in this project. Theoretically, we aim to test the degree to which lexical priming is valid for Mandarin Chinese and whether LPT can account for the range of the language's lexical and syntactical behaviours. Furthermore, we shall attempt to expand LPT's

---

[8]The value will be detailed in Chapter three. Here, we only display the formula as such $MC = (A_t + B_t) * 100 / A_s + B_s$.

dimensions in the field by comparing the lexical and syntactical similarities in and differences between English and Chinese. By examining the evidence presented by the given nodes and the clusters containing those nodes, we aim to assess the potential of applying LPT in cross-linguistic studies of English and Chinese.

Regarding methodology, we shall adopt a data analysis perspective not formulated in "classic" LPT. Specifically, we shall focus on the interaction between collocation, semantic association, and colligations at two or more levels. The range of lexical and syntactical features of both English and Chinese nodes will be examined and then contrasted at a contextual level. The corresponding equivalence, either in the pattern of single node or word sequence, will be selected based on degree of translatability, which will be computed using a database established specifically for the comparators.

2.6 A summary

This chapter has specified the theoretical framework used in the project, and discussed both the literature informing the theory and the methodology adopted. Key concepts examined in the thesis were also mentioned. Our review of studies using LPT as their start point points to the need to further investigate different kinds of language. While this project does not intend to cover all the issues concerning LPT needing further consideration, it will attempt to fill existing gaps by cross-linguistically examining a pair of corresponding lexical items in English and Chinese. Taking keyword analysis as our starting point, we shall conduct phraseology research on word sequences containing the node word to provide a

holistic picture of the lexical and syntactic behaviour of the items. The criteria for choosing the lexical items will be specified in the next chapter, and the methodology used in the project will be detailed.

CHAPTER3

Research Methodology

3.1 Introduction

We discussed in the previous chapter that LPT seeks to explain many linguistic features from a psychological perspective. Primings are a kind of "mental concordance" for a language user that cannot be observed directly in corpus linguistic work. However, one could use a corpus to indicate that certain primings are likely to be shared or avoided by a large number of speakers within a particular community. As Hoey pointed out, a "computer corpus cannot tell us what primings are present for any language user, but it can indicate the kinds of data a language user might encounter in the course of being primed" (Hoey 2005:14). Hoey further explained that, in the study of LPT, a corpus "can serve as a kind of laboratory in which we can test the validity of claims made about priming" (ibid).

This chapter will first specify the procedure of selecting the research words, and then specify the types of corpora and methods used in this study and why, and introduce the analytic techniques used. Additionally, it will specify the methods used to establish a self-built database for particular uses in this study.

3.2 Choosing the research nodes

It is necessary to choose appropriate nodes to reflect a study's purpose. This section specifies and justifies the criteria used to do so in this study.

3.2.1 Four criteria

1. It was decided that the word type to be investigated in this study needed to be
    NOUN. English and Chinese are distinctive languages from many perspectives.
    English has a variety of types of morphological changes for verbs, such as
    agreement markers, tense, and aspect markers; Chinese verbs, however, do not
    manifest a high degree of this kind of grammatical change, even though their
    tense can be deduced from such adverbs as 昨天 *zuo2tian1* (yesterday) and 明
    天 *ming2tian1* (tomorrow) and their aspect from such aspect morphemes as 了 *le*
    (perfective) and 着 *zhe* (durative). While one might obtain surprising and
    interesting findings (e.g., morphological primings) from the comparison of the
    two languages' verbs, those findings are not consistent with the major concern in
    the thesis. In contrast, fewer primings arise with regard to grammatical
    variations among nouns in both English and Chinese, with only the presence or
    absence of plural formation to distinguish them. As Doyle (2003) showed,
    common collocates should never be assumed for various forms of a lemma (as
    shown in Chapter 2). LPT concurs with these linguists and claims that primings
    operate for words and different forms of a lemma could have characteristic
    primings. We shall not, therefore, select a noun with a plural form.

2. The noun chosen for this study should not have too many synonymies and its
    body sense should be unlikely to be affected by cultural differences. In other
    words, the generally recognised meaning of the node nouns should be similar for
    both English and Chinese speakers and should have fixed pronunciation (or
    tone). Thus, words changing their part of speech when the stress shifts
    (e.g., *'process* as noun versus *pro'cess* as verb) were not considered in the study.

3. The nodes to be examined should not have multiple translational alternatives. For example, Xiao's (2006) *result* group has several equivalents in Chinese, such as 后果 *hou4guo3* (outcome) and 恶果 *e4guo3* (aftermath)l thus, lexical and grammatical features revealed by *result* could not be reduced to one equivalence. Because our research focuses on the word *per se* and assumes that every word embraces unique collocational, semantic associational, and colligational features for both English and Chinese, lexical items residing in several patterns of counterparts might complicate our results and fail to meet our initial research goal.

4. As mentioned in Chapter 1, the research nodes need to be used with a similar frequency in both English and Chinese corpora to provide ample data to analyse.

3.2.2 The procedure of choosing the research nodes

Taking these four criteria into consideration, we first searched for word candidates using the website *www.wordfrequency.info/free.asp?s=y*—a free online list of the lemma and parts of speech for the 5,000 most often used words in English. Of the top ten nouns from the list—*year*, *people*, *way*, *man*, *thing*, *woman*, *life*, *child*, *world*, and *state*—only the word *world* met all the criteria set above. First, its Chinese equivalent (世界 *shi4jie4*) was listed as a top noun in the Chinese lemma list (Chinese National Corpus), meaning 世界 *shi4jie4* is generally used as frequently in Chinese as *world* is in English. Second, the body-part senses of *world* and 世界 *shi4jie4* are shared by both English and Chinese speakers. According to the OED, *world* has two major senses: 1. *(usu. the world) the earth, together with all of its countries, peoples, and*

natural features. 2. *a part of aspect of human life or of the natural features of the earth, particular a region or group of countries, human and social interaction, and secular interests and affairs*. These same senses can be found under the entry 世界 *shi4jie4* in the Modern Chinese Dictionary. Additionally, some of *world's* connotational meanings are also shared by its Chinese partner; for example, *bring someone into the world* (把他/她带到这个世上 *ba3 ta1 dai4dao4 zhe4ge4 shi4jie4 shang4*). This high degree of correspondence enables us to conclude that *world* and 世界 *shi4jie4* are largely equivalent. The third and the most important reason for choosing *world* and 世界 *shi4jie4* as the research words is that 世界 *shi4jie4's* compound characters co-select each other and occur frequently with each other, with each representing a similar sense or sub-sense of 世界 *shi4jie4* (where *shi4* means 'one's life in the world' and *jie4* means boundary between two physical worlds and religious worlds). Thus, the polysyllabic form of 世界 *shi4jie4* in Chinese constitutes the single English word *world* in English. This reduces potential confusions that may arise from different word constructions of the two languages and makes the selection of *world* and 世界 *shi4jie4* satisfy the four criteria perfectly.

3.3 The corpora

Three kinds of corpora were used in this study, for different purposes. General corpora were used to observe the general collocational and colligational preference/avoidance of the research nodes. Semantic associations were studied with instances retrieved from the general corpora. The characteristic features discovered in the general corpora were then further proven using corpora containing larger

amounts of data—i.e., reference corpora. We analysed a number of instances in detail, especially when examining the two node words' particular colligational behaviour. As it was impossible to examine every instance containing the node word in detail, due to their highly frequent occurrences (58,496 instances of *world* occurring in BNC and 1,079,563 instances of 世界 *shi4jie4* occurring in zheTenTen 11), a qualitative and manageable annotation was undertaken for the colligational and semantic associational study of the nodes. To complement the analysis, therefore, we built a number of databases comprised of retrieved concordance lines. The procedure for establishing the database of extracted instances is specified in Section 3.3.3.

3.3.1 The general corpora

Two comparable corpora were used in the study to examine general collocational colligational behaviour—the Freiburg-LOB corpus of British English (henceforth FLOB) and the Lancaster Corpus of Mandarin Chinese (henceforth LCMC). "The Freiburg-LOB (FLOB) corpus is a recent update of LOB, which is composed of approximately one million tokens of written British English sample proportionally from fifteen text categories published in the early 1990s" (Hundt et al. 1998, cited in McEnery and Xiao 2007:104). The LCMC was sponsored by the Economic and Social Research Council (ESRC) in the UK and designed by Professor Zhonghua Xiao (Richard Xiao) and his team using the same sampling criteria as for the FLOB; it represents written Mandarin Chinese published in China in the corresponding sampling period (1991-1992). LCMC contains one million words, with approximate 1.6 Chinese characters being calculated as the equivalent to one English word. The

corpus was initially established for cross-linguistic study of English and Chinese. Though numerous English and Chinese corpora had previously been constructed by English and Chinese researchers, respectively, an open and free-access parallel corpus for comparing the two languages was still awaiting development. To fill this gap, Professor Xiao and his team collected 500 texts, each with 2,000 words on average, to establish a comparable corpus to FLOB. LCMC covers similar genres to FLOB, with the exception of Genre N. In FLOB, Genre N comprises texts about Western cowboy stories; however, China does not have this kind of fiction, so this genre was replaced with martial arts novels. The genre types for these two corpora were distributed below, with the different Genre N bolded:

Table 3.1 Genre types of FLOB and LCMC

Type	Register	Code	Genre	Samples	Proportion
Non-Literary	News	A	Press reportage	44	8.8%
		B	Press editorials	27	5.4%
		C	Press reviews	17	3.4%
	General prose	D	Religious writing	17	3.4%
		E	Instructional writing	38	7.6%
		F	Popular lore	44	8.8%
		G	Biographies and essays	77	15.4%
		H	Reports and official documents	30	6.0%
	Academic prose	J	Academic writing	80	16.0%
Literary	Fiction	K	General fiction	29	5.8%
		L	Mystery and detective fiction	24	4.8%
		M	Science fiction	6	1.2%
		N	**Adventure fiction (FLOB) / Martial arts fiction (LCMC)**	29	5.8%
		P	Romantic fiction	29	5.8%
		R	Humour	9	1.8%
Total				500	100%

(adopted from Xiao 2014: 15)

3.3.2 The reference corpora

In addition to the two general corpora introduced above, two reference corpora—the British National Corpus (BNC) and zhTenTen11—were used to complement the data derived from the general corpora, The BNC was created by Oxford University Press in the 1980s and early 1990s. It contains 100 million words taken from texts from a wide range of genres (e.g., spoken, fiction, magazines, newspapers, academia). zhTenTen11 is a simplified modern Chinese corpus containing almost 2.6 million documents with more than 1.7 billion words in over 72 million sentences. BNC and zhTenTen11 were used to test results detected from FLOB and LCMC, due to their larger data capacity, and to retrieve instances of *world* and 世界 *shi4jie4* for detailed annotation and analysis. Through this process, the self-built datasets for this study were established.

3.3.3 Self-built database

The general databases were built by selecting every second concordance line from the general or reference corpora. The number of instances ranged from 100 to 1000. They were first stored in MS Excel, and then analysed manually by the candidate. This dataset type was called the Database of World (DW), referring to instances containing *world* and 世界 *shi4jie4* both. The databases were subsequently ordered DW1-DW15, for particular investigatory purposes that will be described in Chapters 4-7.

The second kind of self-built database was a parallel database, which was built specifically to identify equivalent patterns of word sequences containing the two nodes. As noted in Chapter 2, we did not rely on a bilingual dictionary or the researchers' own language experience for two reasons. First, a word's meaning is realised by collocating with particular words in a particular grammatical pattern (colligation), and cannot be simply reflected in a dictionary entry. Second, as a researcher's or a translator's personal experience may vary from person to person, no one person can authoritatively determine equivalent patterns for cross-linguistic study. Accordingly, we referred to recurrent translational pairs in a parallel corpus (database in the project), based on the different degrees of their equivalence.

We adopted Altenberg's (1999) Mutual Correspondence Value formula, presented below, to help us choose the equivalent pairs:

$$MC = \frac{(A_t + B_t)*100}{A_s + B_s}$$

where A_t and B_t mean the number of times that Word A and Word B are translated mutually, while A_s and B_s are the occurrences of Word A and Word B, respectively, in the source language texts. Assigned values for this formula were obtained as presented in Figure 3.1, below:

Figure 3.1 The process of obtaining values for MC formula
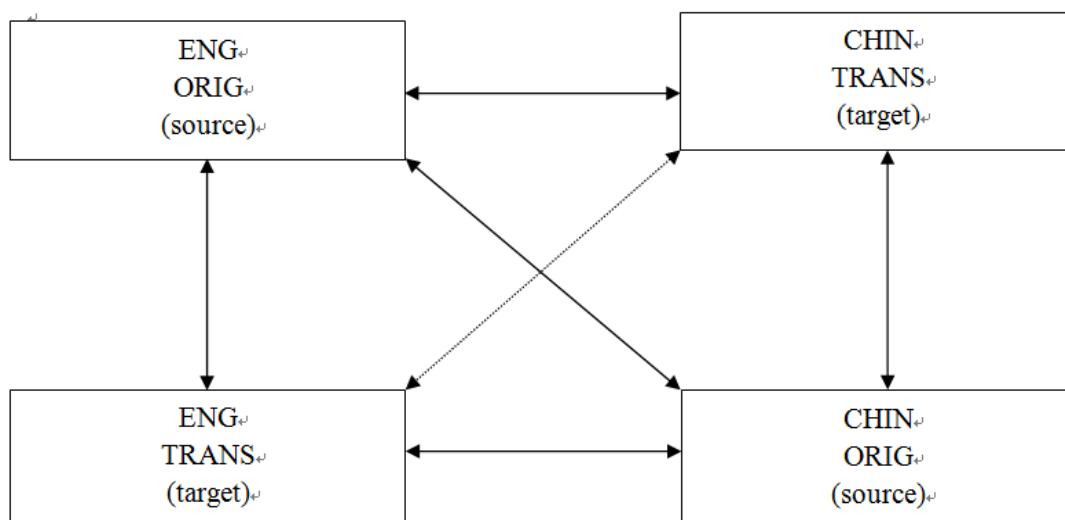
<div align="right">Johansson and Hofland (1994: 26)</div>

To specify the procedure for building up the parallel database using the above figure, we used the *result* word group to detail the step in each blank shown above:

- Instances containing *result, consequence, aftermath*, etc. were retrieved from FLOB, thus accumulating the ENG ORIG texts (Figure 3.1, top left corner).

- These instances were first translated into Chinese by the candidate and then revised by two colleagues[9]. The database for CHIN TRANS texts (top right corner in Figure 3.1) was thus formulated.

- The source (ENG ORIG) texts containing *result* were placed into the left column of MS-Excel and the corresponding translated version were placed into the right column, as shown below:

---

[9] The two colleagues are associate professors at the English Department of Dalian University of Foreign Languages and are experience translators and teachers of English.

The result did not take into account yesterday's Gallup poll, which gave the Tories a 4.5 per cent lead, or the latest fall in interest rates.	此项结果并没有将昨天的盖勒普民意调查以及最近的汇率下降计算在内，其中，盖勒普民意调查显示托利党比对手多赢得4.5%的选票
After a brilliant decade for investment performance during the 1980s, UK pension fund managers were overdue for a market setback. In fact, the minus 10.5 per cent median rate of return was the first negative result in nominal terms since 1974.	在上世纪80年代经历了良好的投资回馈之后，由于市场萧条，英国养老基金领导人显得过多。事实上，自1974年起，10.5%养老金员回报率已经成为了最槽糕的结果。
The result rescued a tour that was floundering after just one win in four attempts.	在经历了四局比赛只有一局胜出的艰难长时候，该结果挽救了整场比赛。
It also raises the possibility that the council may be poll tax capped as a result of the massive deficit.	该会议极有可能在大量财政赤字的前提下对税收进行民意调查。
The Seahawks were beaten 9-6 at Telford on Tuesday but the BIHA overturned the result and gave the Seahawks a 5-0 win after Telford illegally iced import Dan Sweeney.	在周二，Seahawks以9比6的比分完胜Telford，但是在Telford以不正当形式搁浅买进外援Dan Sweeney之后，BIHA对逆转局面，以5比0完胜Seahawks.
A Home Office spokeswoman said that the disciplinary inquiry now under way would consider whether anyone should face action as a result of failure to pass that information on to senior Prison Department and Home Office staff and Ministers.	内政部发言人称纪律审查程序包括是否对未能向高级监狱部门，内政部工作人员以及部长上报有关信息的人提起诉讼。
That doesn't mean we'll beat France but we may achieve a result that not many are expecting.	这并不意味着我们能战胜法国队，但是我们也许会获得意想不到的结果。
The result was that almost every shot was hit flush out of the middle of the club.	结果是几乎每个坐在俱乐部中间的人都抓到了同花顺。

- The candidate Chinese equivalents to *result—结果 jie2guo3 (result), 后果 hou4guo3 (consequence/outcome), 恶果 e4guo3 (aftermath)*, etc.—were retrieved from LCMC and accumulated as the CHIN ORIG texts (at the right bottom corner in Figure 3.1).

- Instances containing *结果 jie2guo3 (result)* were first translated by the candidate and her colleagues, then revised and checked by a native speaker from Britain[10]. Cases of *结果 jie2guo3* were put into the left column of MS-Excel and the translated version into the right column, as shown below, thus forming the CHIN ORIG texts (the left bottom corner in Figure 3.1):

---

[10] A lecturer and also my dear friend Aaron Roger — a native English speaker and is currently working at Ningbo Nottingham University.

结果_n ，_wd 不知_v 是_vshi 长虹_nz 厂_n 领导_n 说_v 对了_vl ，_wd 还是_c 别的_rzv 什么_ry 原因_n ，_wd 他们_rr 这_rzv 次_qv ″_wy 犯上_v 决策_vn ″_wy 没有_d 受到_v 制裁_vn 。_ew </s>	However, they were not punished for making this decision. We don't know what reason should be accounted. Maybe the head of Changhong Factory was right or maybe other reasons.
调查_v 结果_n 表明_v ，_wd ７６％_m 的_ude1 企业_n 对_p 中国_ns 员工_n 的_ude1 素质_n 表示_v 满意_v 。_ew </s>	The research result shows that 76% foreign companies are satisfied with Chinese Staff's working spirit.
《_wkz </s> <s n="029"> 长春市_ns 城市_n 规划_n 管理_vn 条例_n 》_wky （_wkz 修改稿_n ）_wky 经过_p 一_m 下午_t 的_ude1 审议_vn 还是_c 没有_v 结果_n 。	There is still no mutual agreement on the deliberation of Regula
结果_n 除了_p 两_m 位_q 外出_vi 探亲_vi 、_wn 一_m 位_q 妻_ng 生_v 重病_n 请假_vi 外_f ，_wd １０_m 个_q 党员_n 全部_m 到_v 齐_a 。	Surprisingly, every member of the Communist Party arrived except three members (two had been out of the town to visit his relatives and one had to take care of his wife who was seriously ill.
比赛_vn 结果_n ，_wd 中国_ns 国家队_n 和_cc 捷克斯洛伐克_nsf 布拉格_nsf 斯_b 巴_b 达_v 队_n 同_p 以_p １：０_m 的_ude1 比分_n ，_wd 分别_d 战胜_v 了_ule 香港_ns 钻_v 禧_ng 队_n 和_cc 苏联_ns 莫斯科_nsf 斯_b 巴_b 达_v 队_n 。_ew </s>	The match result was that China and Czechoslovakia were beaten 1-0 at Hong Kong and Russia respectively.
另_rz 一_m 场_qv 女单_n 半决赛_n 的_ude1 结果_n 是_vshi ，_wd 黄华_nr ２：１_m （_wkz １１： １２_m ，_wd １２：１１_m 、_wn １１：４_m ）_wky 胜_v 队友_n 唐九红_nr 。_ew </s>	Another semifinal result was 2-1, with Huang hua winning over her team member Tang Jiuhong.
国际_n 舆论_n 认为_v ，_wd 黎_b 叙_b 两_m 国_n 总统_n 的_ude1 会晤_vn 结果_n ，_wd 是_vshi 对_p 奥恩_nrf 为首_vi 的_ude1 基督教_nz 强硬派_n 以及_cc 主张_v 黎巴嫩_nsf 实行_v ″_wy 邦联_n ″_wy 计划_n 的_ude1 一个_mq 打击_vn 。_ew </s>	The global press believe that the meeting between the two presidents of The Republicof Lebanon and Syria was a strong hit on Christian hardliners headed by Michel Aoun and leaders who support federation in Lebanon.

The procedure of building this parallel database is summarised in Figure 3.2:

Figure 3.2: An example of comparing patterns in the self-built parallel corpus
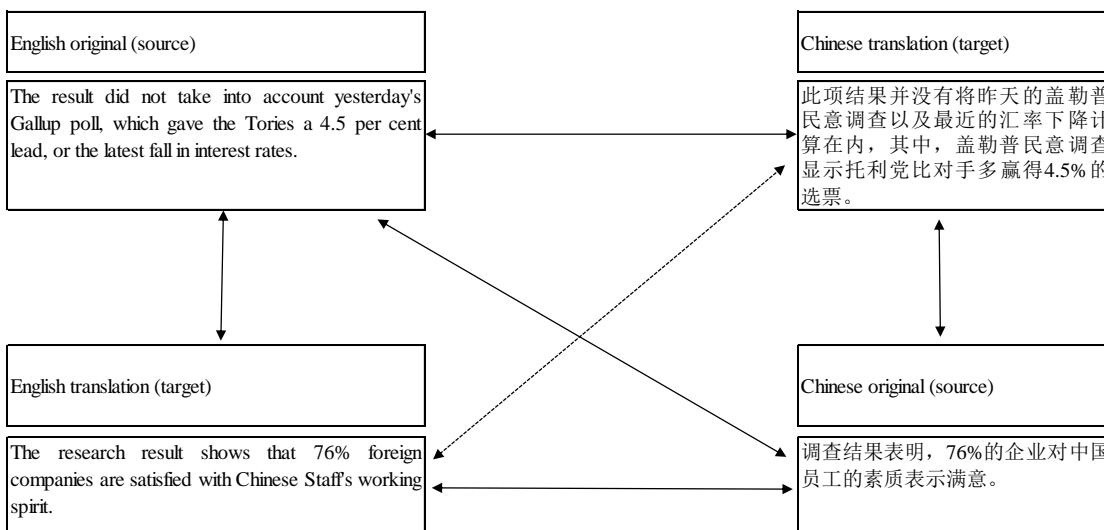
Unlike other parallel corpora, which have been built through unidirectional translation (i.e., from a source language to a target language; e.g., from English to Chinese), the bi-directional structure used in this project compared the translated version to the linguistic features of the native language (i.e., compared translated

English with native Chinese) and vice versa. The high quality of the English and Chinese texts as both source and target languages, with respect to MC value, enabled us to choose the best equivalent pairs containing *world* and 世界 *shi4jie4* (world).

3.4 Instrumentations

A number of corpus analytic tools and software developed by linguistic and computer science researchers enable computers to manipulate database contents. Some analytic tools are designed for a specific corpus while others are open platforms capable of analysing any text or corpus uploaded; both can be run on researchers' personal PCs over the Internet. Unless specialised tools are applied to observe and analyse them, corpora are just large accumulations of data. Hence, analysing tools are becoming important factors for corpus linguistics when dealing with complex statistical data.

In this study, two corpora analysing tools were used. One was AntConc (3.2.4w), created at Waseda University by Dr. Laurence Anthony, a much honoured physicist. AntConc is a stand-alone tool that began as a simple concordance program, but gradually progressed and became a powerful text analysis tool providing such functions as Concordance, Word Clusters, Collocates, Keyword list, etc. An important and unique function of AntConc is its ability to analyse Chinese texts in its UTF8 version. However, the version provided by LCMC on its home page uses .txt format, which is not compatible with the format required by AntConc. These .txt documents were therefore converted into UFT8 version using the UTF Convertor and then analysed with AntConc.

Another software applied in this study was the Sketch Engine, created by Adam Kilgarriff (Kilgarriff *et al*., 2004). It is a Corpus Query System with a wide range of functions, such as "concordancing, word sketches, and a distributional thesaurus" (Kilgarriff et. al., 2015:63). Word Sketch is a distinctive feature of this software. Like other corpora tools, it can list words that commonly collocate with the node word; however, it can also present their grammatical relationship with the keyword. Figure 3.3 shows that the most frequent verb collocating with *result* (as Object) is *achieve* and the most frequent verb collocating with *result* (as Subject) is *indicate*.

Figure 3.3 Word Sketch output for *result*

verbs with "result" as object			verbs with "result" as subject			"result" and/or …		
	8,838	0.26		6,658	0.20		3,025	0.09
achieve ✚	314	9.16	indicate ✚	126	8.49	test	23	7.37
obtain ✚	277	9.05	results indicate that			performance	15	6.64
results obtained			show ✚	285	8.48	conclusion	10	6.52
produce ✚	490	9.04	results show that			number	23	6.46
publish ✚	151	8.23	suggest ✚	179	8.43	. as a result , a number of		
report ✚	137	8.16	results suggest that			report	17	6.46
yield	88	8.16	confirm	69	7.83	action	16	6.45
interpret	76	7.86	results confirm			cause	11	6.44
desire	67	7.78	demonstrate	40	7.24	study	15	6.27
the desired result .			results demonstrate that			method	11	6.05
announce	93	7.70	support	43	6.90	background	8	6.03
show ✚	207	7.65	these results support			score	7	6.02
compare	88	7.64	obtain	26	6.69	datum	9	6.00
analyse	56	7.35	the results obtained			experiment	7	5.95
present	83	7.30	be ✚	4,020	6.57	theory	10	5.92
express	76	7.20	result is			steel	7	5.89
results are expressed as			reflect	26	6.44	asset	7	5.89
await	45	7.15	results reflect the			dividend	6	5.89
confirm	49	7.01	encourage	21	6.30	cent	7	5.86
expect	76	6.91	present	22	6.13	finding	6	5.81
get ✚	312	6.87	the results presented			process	10	5.78
give ✚	300	6.83	follow	48	5.94	detail	8	5.76
predict	34	6.71	results were as follows :			technique	8	5.75

This function was extremely helpful for analysing words' semantic association in this

study, as it saved the researchers from having to read every line of the concordance to group collocates into different semantic associations. A page like that shown above told them everything and reduced the time and effort they had to expend.

By clicking on the number next to the word *achieve*, one can see all instances containing the combination of *achieve … result* (see Figure 3.4).

Figure 3.4 Concordance for *achieve … result*

over adversity and *achieved* such tremendous results this season and I am sure that members
be called the cost-effectiveness of the results *achieved* by Bomber Command. The C-in-C
any fund, but they *achieved* the same end result . Jimmy? Well if this is the case then you
and the statute so understood *achieves* no results not intended by all those who voted for
commercially necessary to *achieve* an optimum result . </p> Conditions precedent <p> Where there
suggestion that the reference, to *achieve* those results , would have had to have been to arbitration
included in a contract to *achieve* the same result ; for instance: It is agreed that if any
prohibited steps order with a view to *achieving* a result which could be achieved by a residence
<p> Once you know that you have *achieved* a result then obtaining your costs from the opposition
that if she did not *achieve* the necessary results for entry to college in the Summer of ninetee
Both managers and engineers *achieve* their results through other people. Both are obliged
Local Centres </p><p> Following the excellent results *achieved* in 1992 (61% pass rate), Lloyds
80 per cent higher than 1991. </p><p> These results were *achieved* in spite of very competitive
helped the company to *achieve* its 1992 record results . Left to right, , Albion Maltings, Ipswich
feed costs. Fulmar *achieved* a very good result in 1992, with profits ahead of expectation
necessary for a successful GIS. Display of results is *achieved* in one of two forms: hardcopy
system these sentences had *achieved* very poor results whether as a result of the grammatical
Glasgow, and much home work, *achieved* good results in his finals for the Institute of Management
searching at Aphrodisias itself. </p><p> The results *achieved* in 28 summers of work have been
respectively, but hopes to *achieve* even better results by using the technology developed through

Sketch Engine is also a professional character-based Chinese analysis system that can investigate features of both single and multiple Chinese characters. Sketch Engine can generate concordance and frequency lists for all words that incorporate a

particular character, such as the 果 *guo3* in 结果 *jie2guo3* (result), enabling the user to find other words incorporating the character 果 *guo3* and view their frequency.

Another unique function is Sketch Diff, which can "help learners understand how similar words differ in order to help with the choice of the right words." (Kilgarriff 2015: 71). Figure 3.5 below shows the Sketch Diff results for two similar words belonging to the *result* group: *outcome* and *consequence*. The words highlighted in green are more likely to collocate with *consequence*, while those highlighted in red are more likely to collocate with *outcome*. White words collocate equally with both.

Figure 3.5 Sketch Diff for *outcome* and *consequence*

consequence	6.0	4.0	2.0	0	-2.0	-4.0	-6.0	outcome

"%w" and/or ...	904	501	0.12	0.11
antecedent	22	0	9.6	--
cause	40	0	9.3	--
determinant	6	0	7.7	--
implication	9	0	7.5	--
behaviour	8	0	6.4	--
illness	5	0	6.4	--
nature	8	0	6.2	--
term	5	0	5.5	--
event	5	0	5.5	--
action	5	0	5.4	--
condition	6	0	5.3	--
cost	6	0	5.2	--
level	5	0	5.0	--
change	5	0	5.0	--
course	8	5	5.2	4.6
use	5	5	5.2	5.4
outcome	5	6	6.9	7.6
quality	0	5	--	5.5
care	0	5	--	5.5
decision	0	5	--	6.1
procedure	0	6	--	6.3
consequence	0	5	--	6.9

verbs with "%w" as subject	548	678	0.07	0.15
arise	7	0	5.3	--
occur	8	0	5.2	--
follow	20	0	5.1	--
have	55	47	2.1	1.9
do	5	5	1.0	1.0
be	337	481	3.0	3.5
seem	0	9	--	3.9
reflect	0	6	--	5.7
depend	0	11	--	6.2

Highlighting a keyword's collocates in different colours enables researchers to spot

major collocational differences between the two words quickly, while indicating node words' grammatical function (e.g., as subject or as object) facilitates comparing the two words' syntactical features at the grammatical level.

In addition to the large capacity of corpora of different languages, Sketch Engines' new but very reasonable statistic association score (LogDice) gives very good results for collocation candidates in a large corpus. Compared with MI (Mutual Information value) scores, which to a large degree are decided by the product of the node word and its collocating candidates, LogDice focuses on the co-selecting strength of the node word and its collocates. MI score might be determined by the high frequency of a particular word in a particular corpus; however, LogDice is arguably "more reliable since it will not biased by either too high or too low frequency of the items in the query" (Rychlý, 2008 cited in Hoey and Shao 2015:23).

LogDice scores do not always correspond to the actual number of occurrences of the apparent collocates. To avoid data misinterpretation in subsequent chapters, we explain this non-correspondence in Figure 3.6 below.

Figure 3.6 Distribution of L1 collocates of *result* in descending LogDice order

	Cooccurrence count	Candidate count	T-score	MI	logDice
P | N The	3,315	619,511	53.440	3.799	7.356
P | N election	258	8,695	15.854	6.270	7.348
P | N These	316	26,353	17.206	4.963	7.218
P | N direct	238	11,028	15.152	5.811	7.168
P | N a	8,729	2,041,849	85.030	3.475	7.099
P | N end	277	46,230	15.575	3.962	6.665
P | N examination	126	4,395	11.074	6.221	6.439
P | N positive	131	8,042	11.175	5.405	6.389
P | N may	373	111,435	17.095	3.122	6.304
P | N Our	126	9,195	10.910	5.156	6.300
P | N test	129	11,049	10.983	4.925	6.284
P | N final	129	14,646	10.862	4.518	6.191
P | N similar	128	17,339	10.724	4.263	6.115
P | N net	102	5,915	9.874	5.487	6.089
P | N good	227	73,095	13.202	3.014	5.999
P | N best	148	33,176	11.117	3.537	5.989
P | N This	295	110,097	14.712	2.801	5.978
P | N better	142	33,473	10.836	3.464	5.923
P | N these	241	97,128	13.119	2.690	5.814
P | N the	9,255	5,415,707	74.568	2.152	5.795
P | N same	163	60,650	10.941	2.805	5.684
P | N will	451	245,313	16.797	2.258	5.678
P | N exam	66	839	8.084	7.677	5.618
P | N which	602	361,203	18.878	2.116	5.608
P | N has	429	254,360	15.992	2.133	5.562

A good example is the word *the*, which has been shown to be an inseparable component of the noun phrase (Stubbs 2007, Hunston 2001), as shown by its high frequency of appearance in the whole corpus; indeed, in the entire English language, based on its appearance as the most frequently used word in almost every widely-used general corpus (BNC, BoE, etc.). Yet, it ranks twentieth in the L1 collocation list above, which implies *the* is not the strongest collocate to the node word *result*, even if it is one of the most frequently-used words in English.

When we put *world* into the query entry and order its left collocates in a descending MI score, we discovered that the top five immediate collocates in L1 (first place to the left of the node word) position for *world* were *post-Second*, *Starcoast*, *pre-First*,

post-First, and *Non-Western*. Based on our common sense, these words tend to be more rarely used with *world* than are such collocates as *third*, *second* and *first*. The reason that for this seemingly wrong picture was that the MI score was treating those lower-frequency words as strong collocates; *post-Second,* for example, only occurred 12 times in BNC, but always with *world*, so MI calculated it to be the strongest collocate thereof. While this might offer insight into the word *post-Second*, is less obviously insightful about *world* and does not meet the aims of this study.

Another way of calculating a word's collocational strength is the T score, described by Hunston (2001: 16) as a value "measure[ing] the degree of certainty with which one can say that co-occurrence is due to the behaviour of the target item rather than to chance." This measurement tends to produce a very different picture from that of MI. In terms of the top five T-scores, the immediate collocates for *world* in L1 position were *the*, *third*, *second*, *first* and *real*. These results are much closer to the results derived from the LogDice scores, but do not necessarily contribute to the actual collocational profile of the node word, because the score gives too much weight to such high-frequency words as *the*. Both aforementioned scores are inappropriate for revealing the objective collocational strength of a node word. LogDice, on the other hand, can solve this problem by adding a log algorithm to the formulae.

3.5 The research design

After arguing that the claim that the information yielded by corpora is based more towards performance than competence is overstated, Leech (1992: 108) concluded

that the distinction between competence and performance is not as great as claimed, "since the latter is the product of the former." Consequently, "what one discovers in a corpus can be used as the basis for whatever theoretical issue one is exploring" (Meyer 2002: 4). The research design for this study is based mainly on, and uses language "performance" to test and modify, Hoey's paradigm for researching his priming hypotheses. The theoretical framework adopted in this thesis is shown below:

Figure 3.7 The schematic representation of the theoretical framework
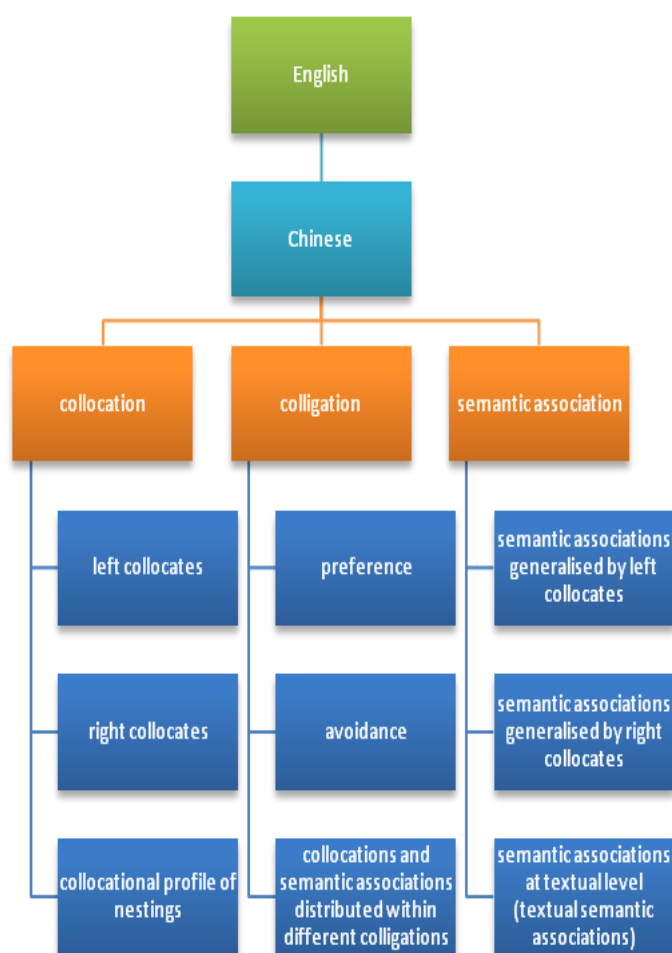
The steps taken in the research process of this study and how these steps were carried

out to answer each research question are illustrated as follows:

1. First, we examined the general collocational behaviour of *world*. The collocates were investigated based on the positions in which they occurred. We then investigated *world*'s general colligational behaviour. We studied *world* occurring at different grammatical positions within different grammatical functions (Subject, Object, Complement and Adjunct) to collect data for comparison with its Chinese equivalent 世界 *shi4jie4* (world).

2. Next, we explored *world*'s collocation and semantic associations in combination with its colligational behaviour. Complex issues concerning the collocational and colligational behaviour of nestings and semantic associations (in a broad sense) were also considered.

3. To test the validity and applicability of LPT for Chinese, we examined the collocational, colligational and semantic associational behaviour of 世界 *shi4jie4* (world), using the procedure outlined for *world* (see Chapters 6 and 7).

4. Translational pairs in this thesis were decided based on MC value, as detailed in Section 3.3.3. The parallel corpus was of a small size, as it was established only for the research purpose of this study.

The following chapter reviews the general collocational and colligational behaviour of *world* to answer partially the question, "What characteristic similarities and differences do English and Chinese display in respect of their lexical behaviour?"

.

CHAPTER 4

Collocations and Colligations of *world*

4.1 Introduction

This chapter provides evidence of how collocation operates for the English word, *world*. We started our research from this point for two reasons. First, the collocational behaviour found for the node word *world* can extend and advance LPT with evidence found for English. Previous linguists have either restricted their studies to a single genre[11] (e.g., Hoey retrieved instances from the *Guardian* corpus) or applied corpora of relatively smaller size (e.g., Pace-Sigge retrieved instances from SCO, which only contains 120,000 words). To supplement the data, larger corpora of various types, including the FLOB and LCMC as comparable corpora, BNC and zhTenTen11 as general corpora, and a self-built parallel database were used in this study. Second, evidence found for the collocational, colligational, and semantic associational features of *world* could be referenced to partially answer the second research question, "What characteristic similarities and differences do English and Chinese display in respect of their lexical behaviour?" Before comparing the characteristic similarities and differences between *world* and 世界 *shi4jie4*, we had to know the lexical behaviours of the two languages.

Since the property of collocation is the starting point for LPT, examining this property

---

[11] Hoey argued that all the claims proposed in the Lexical Priming Theory were in the first place constrained by domain or genre. We agree with this claim in that languages may perform differently in specific situations and contexts. What we intend to claim here is that corpus used by Hoey is limited in genre and size.

was a reasonable means of finding out whether the theory could be applied to Chinese. Our strategy was to first examine collocation, colligation, and semantic association for the English word *world*, and then see whether the same behaviours occurred for the Chinese equivalent 世界 *shi4jie*. Collocation is defined herein as:

> …a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution. (Hoey 2005: 5).

4.2 The general collocational profile of *world*

4.2.1 The left collocates of *world*

We began by retrieving a concordance of *world* with a span of five (±5) tokens to the left and right sides. A total of 58,496 instances were found in the BNC corpus, with 520.94 occurrences per million running words. Analysing these 58,496 instances in Sketch Engine, we found a greater consistency of patterning and sets of variables to the left of the collocation than to its right.

The collocates of the node word *world* were displayed in descending order sorted by LogDice scores, after which an outline of collocates within the span of ±5 tokens was displayed. L1, L2, and L3 were used to refer to the first, the second and the third position to the left side of the node word; the same convention applied to the right side (e.g., R1, R2 and R3).

Table 4.1 below shows that the L1 position of *world* was occupied frequently by the words, *Third*, *Second* and *First*, with the "inherent component" (Sinclair 2004:31) *the*

ranking in eighth place. As explained in Chapter 3, the higher frequency of *the* preceding to a noun phrase might be attributed to the frequency of use of the word itself; therefore, *the* did not appear as the strongest left collocate of the node word *world* when data were sorted using LogDice scores.

Table 4.1 Collocates of *world* at L1 position (sorted by LogDice score)

Rank	Collocate	LogDice
1	Third	9.616
2	Second	9.503
3	First	8.832
4	real	8.071
5	outside	8.032
6	Arab	8.025
7	whole	7.311
8	the	7.294
9	modern	7.099
10	New	6.781
11	third	6.677
12	Western	6.670
13	natural	6.630
14	developing	6.572
15	second	6.548

Another interesting finding was that collocates for *world* occurring at the L1 position tended to start with a capital letter, such as *Third*, *Second*, *First*, *Arab*, *New* and *Western*. This is a typical collocational behaviour for *world*, as not every noun in English is preceded by a premodifier starting with a capital letter.

The L2 position to *world* was predominantly occupied by prepositions. Table 4.2 below displays the top ten L2 collocates of *world*, all of which were all prepositional

words. This significant colligational behaviour of *world* will in detail in Section 4.4.

Table 4.2 Collocates of *world* at L2 position (sorted by LogDice score)

Rank	Collocate	LogDice
1	around	9.616
2	throughout	9.503
3	outside	8.832
4	over	8.071
5	in	8.032
6	of	8.025
7	round	7.311
8	across	7.294
9	into	7.099
10	during	6.781

Table 4.3 (see next page) shows that collocates at the L3 position of *world* were fewer in number, with only seven collocates occurring more often than random distribution. Four words seem to be a proper sequence number (Sinclair 1991) for discovering a collocation, while the co-occurrence of a word sequence of more than four words may not be more significant than a random distribution. This claim was echoed in the examination of the R3 position for the node word *world*, wherein no collocate occurred significantly more frequently than others.

It was immediately clear that two superlative words—*best* and *largest*—occupied top positions to the right side of *world* (in Tables 4.4 and 4.5), with *largest* ranking 6th and *biggest* ranking 13th in the R1 and R2 positions, respectively.

Table 4.3 Collocates of *world* at L3 position (sorted by LogDice score)

Rank	Collocate	LogDice
1	parts	7.853
2	rest	7.793
3	all	6.905
4	anywhere	6.708
5	countries	6.339
6	best	6.304
7	end	6.125
8	largest	6.033

4.2.2 Right collocates of *world*

We noted that, as with L1 collocates, R1 collocates also had a tendency to start with a capital letter (*Cup*, *Bank*, *Champion*, *Health*, *Broadcast*, etc.). The node word *world* in English typically collocates with words at the R1 position to form a fixed phrase, such as *world war* (28.67 per million in BNC), *World Cup* (10.83 per million in BNC), *World Bank* (4.99 per million in BNC), *world champion* (4.28 per million in BNC), and so on. These combinations (or nestings in LPT) in turn have their own collocates and may be associated with a different semantic meaning than or sometimes share semantic association with its components. This complexity will be investigated further by combining its colligational behaviours in Chapter 5.

Table 4.4 Collocates of *world* at R1 position (sorted by LogDice score)

Rank	Collocate	LogDice
1	war	10.706
2	Cup	9.998
3	Bank	8.648
4	Champion	7.560
5	war	7.478
6	largest	7.360
7	trade	6.886
8	countries	6.876
9	Health	6.766
10	economy	6.702
11	Broadcasts	6.683
12	title	6.651
13	market	6.693
14	champions	6.585
15	championship	6.477

Compared with collocates appearing to the left side of the node word *world*, right-side collocates seem to have fewer variables, particularly at the R2 position, where the most frequent collocate was the Roman numeral *II*. Together with the top R1 collocate (*war*), *II* constitutes a frequently occurring combination—*world war II*.

Table 4.5 Collocates of *world* at R2 position (sorted by LogDice score)

Rank	Collocate	LogDice
1	II	10.706
2	title	9.998
3	countries	8.648
4	Championship	7.560
5	market	7.478
6	championship	7.360
7	economy	6.886
8	championships	6.876
9	champions	6.766
10	where	6.702
11	biggest	6.693

12	most	6.651
13	which	6.585
14	championship	6.477

Of the fourteen most frequent R2 collocates, four were versions of *champion*, including *Championship* (4th), *championship* (6th), *championships* (8th) and *champions* (9th).

Apart from *champion* group words, *world* is also followed by such words as *biggest* (6.693) and *most* (6.651). This implies a strong colligational behaviour of occurring with superlatives, which in turn suggests English speakers have a strong tendency to use *world* as a premodifier serving an evaluative function. Because *world* is the largest physical surrounding for creatures living on earth, it is logical for individuals to use this word with superlatives to emphasise the degree of a given descriptor.

Our examination now shifts from collocation to colligation, because of the dominant occurrences of prepositions at the L2 position and superlatives at the L3, R1 and R2 positions. Due to their grammatical functions, these candidates should be categorised not only as a set of collocations but as instances of colligation. The following section, considers the colligational behaviour of *world* in the corpus without regard to collocation, and then analyses the data by combining collocations and colligations.

4.3 A colligational description of *world*

Colligation must be analysed and tagged manually, as there were no tools to perform this labour. As mentioned in Section 4.2, 58,496 concordance lines of *world* were

found in the BNC, making analysing every line beyond the practical scope of this thesis. Thus, we turned to a smaller corpus—the FLOB—to retrieve a representative but manageable sample of instances with which to investigate the potential colligations of the node word *world*.

A total of 735 concordance lines containing *world* were found across different genres in the FLOB, half of which were then extracted for an examination of colligational behaviour, as follows:

1. First, we extracted half of the instances from each genre (from Genre A to R) by selecting every second concordance line. Instances occurring as the title of a text or that were not part of a complete clause were excluded, and the next valid instance selected instead.

2. Then we copied the selected instances from each text and pasted them into MS Excel. Each concordance line was analysed and tagged manually by consulting their contexts, displayed in Antconc.

4.3.1 An outline of the colligational behaviour of *world* in respect to the verbal groups that occur in the same clause

We extracted 365 instances for our analysis. We first examined the distribution of *world* in clauses across the primary tenses—past, present and future. It has been argued by some linguists that future tense in English might not reflect a real action that will be conducted in the future; some linguists (e.g., Halliday, 2007) even claim that future tense should not be considered a tense at all. However, the present study

adopted the traditional view on future tense because instances belonging to this tense group are few (only five of the 365 instances in our data). The definition of future tense adopted herein is that cited in Collins dictionary: "In grammar, the future tense of a verb is the one used to talk about things that are going to happen. In English, this applies to verb groups consisting of 'will' or 'shall' and the base form of a verb. The future perfect tense of a verb is used to talk about things that will have happened at some time in the future." (online Collins Dictionary, accessed May 11th 2016). Following this definition, we categorised instances with words such as *will* and *shall* as future tense; for instance, *The city's airport will become the world's largest* was treated as belonging to the future tense group in our study, even though some may argue that *will* behaves like a modal verb in this sentence. The structure *be going to* was also treated as a member of the future tense group.

According to Halliday (1993: 36), primary tense is the "deictic component in the tense system, and hence is entered only by finite clauses…". We therefore restricted the investigation at this stage to finite clauses. Clauses whose deixis is achieved by modality (excluding *will* and *shall*, which were treated as markers of future tense) were not included in this research. Since there were only five instances containing modality in the sample corpus, we excluded these examples from our analysis.

Second, we examined the instances from the perspective of voice—active vs. passive. We confined our discussion to the structure of be/get + past particle, which is considered the "norm for English passives" (Xiao 2015).

Aspect—i.e., perfect vs. non-perfect—was investigated at the third stage. Typical past

participle *has/have/had done* were looked for and used to distinguish between perfect and non-perfect.

The procedure for counting these three kinds of colligational behaviour was similar to Halliday's (1993) work, in which he undertook a quantitative study to test the hypothesis that: "general grammatical systems would not be distributed evenly across the probability scale, with all values from 0.5 : 0.5 to 0.99 : 0.01" (ibid: 35). For example, Halliday counted total figures and presented the following percentages: past 598,065 (50.02%); present 597,645 (49.98%).

Halliday's research is concerned with English as a whole, and the corpus he used was much larger than that which we used, so we needed to work out a "normalised distribution" specific to the corpus from which our instances were extracted. If the distributions of the sample corpus made up by extracted instances containing *world* corresponded to the normalised distributions in the FLOB, it would indicate there was no need to discuss the tense profile of *world* in detail, because it conforms to the general tendency of the whole corpus. If, in contrast, the sample corpus' distribution were different from the FLOB's normalized distributions, the issue may merit further exploration.

The procedure for calculating the normalised distribution of each tense in the FLOB was similar to that used by Halliday in his research, and is explicated as follows (We will also calculate the normalised distribution of each 'tense'[12] for the LCMC corpus

---

[12] In Chinese grammar, there is no terminology of "tense" like English. But this does not mean that Chinese does not have a tense semantically. This is realised through markers such as 了 *le1*, 的 *de*, 会 *hui4* and so on. The discussion on Chinese 'tense' will be explicated in Chapter 7.

in Chapter 7 for Chinese *世界 shi4jie*):

- Step one: We selected fifty percent of the texts from the FLOB; specifically, we chose every second text from each genre. For example, Genre A had 44 texts in total, so we chose 22 texts from this genre for the calculation.

- Step two: We identified and counted all finite clauses having modal deixis in the selected texts and removed these instances from the dataset.

- Step three: Within the set of finite clauses remaining (41747 instances), we identified and counted those with verbal groups whose primary tense was Future (as earlier defined).

- Step four: We labelled the remaining set non-future and divided it based on whether the primary tense was past or present.

- Step five: We calculated the percentage of future, present and past within the total set of primary tense clauses; for example, the normalised distribution of present tense in the the FLOB was 55% (23059/41747).

At step one, we obtained 259 texts as a database for calculating the normalised tense distribution of the FLOB. At step two and step three, we have 41741 finite clauses for the identification and counting of tenses. Of these clauses, 1252 were recognised as having verbal groups with future tense by applying the criteria set in Collins Dictionary. In the remaining 40489 instances, which were also identified as non-future clauses, there were 18210 cases that were identified as occurring with present tense verbal groups and 23059 cases were identified as occurring with past tense verbal groups. Finally, we calculate the normalised tense distribution of the FLOB in the formula of OCCURRENCES OF TENSE/TOTAL NUMBER OF

CLAUSES. In doing so, we find that the present tense distribution of the FLOB is 55%, past tense distribution is 42% and future tense distribution is 3%.

Tense distributions of instances containing *world* in the sample corpus were calculated with the same method, and the comparison of the distribution of tense in clauses containing *world* and in the FLOB as a whole is shown in Table 4.6.

Table 4.6 The Comparison of tense distributions between *world* and normalised distributions of the FLOB

Tense Distributions of clauses containing *world*		Distributions of Tense Across FLOB	
Past	34% (124 / 365)	Past	42% (18210 / 41741)
Present	65% (237 / 365)	Present	55% (23059 / 41741)
Future	1% (5 / 365)	Future	3% (1252 / 41741)

The ratio between Past and Present in the FLOB and in the corpus used by Halliday in his research are similar, with the former being 42:55 and the latter around 50:50. This finding can be interpreted in two ways. First, it corresponds to Halliday's research results on primary tense (Halliday 1993:55)—regardless of the inaccuracies he mentioned, such as "extrapolations from samples" (ibid: 55)—because it was unlikely to significantly affect the figures he had gained. The data in this study were obtained from a corpus of around one million words, which is almost 18 times smaller than that used in Halliday's quantitative study. Compared with Halliday's COBUILD corpus, the FLOB was a "sample," with its own tense-use tendencies; specifically, the FLOB preferred present tense slightly more than past tense.

To test whether our interpretation of the distributions from our sample corpus and the FLOB were correct, we then conducted a T-test, using SPSS. We set the tense distributions of the FLOB as the test variable and compared them with the tense distributions of the sample corpus containing *world*. The test result showed a sig. value (p value) of < 0.01, meaning a significant difference existed between tense

distributions in the sample and FLOB corpora. We then set the tense distributions from Halliday's research as the test variable and compared them with the tense distributions of the FLOB; the T-test showed a strong significance ($p < 0.01$).

According to the formula set in SPSS, if the p value is lower than or equal to 0.01, the hypothesis (H_0) that there is no significant difference between sample and test variables is refuted; i.e., a significant difference exists between the sample and test variables.

The T-test results showed a significant difference in tense distributions between the sample corpus and the FLOB, and between the FLOB and Halliday's corpus. We can safely conclude, based thereon, that *world* does have a strong preference for occurring in present tense clauses. This could be somewhat attributed to the FLOB's containing more present tense clauses than general English does, but not decisively so.

With regard to the distributions between passive and active voice, a significant difference was discovered. Our data (in Chart 1 below) showed that only 2% (8/365) of instances containing *world* were used in clauses with verbal groups in the passive voice, while 98% (357/365) were used with active voice, for a passive-to-active ratio of 1:49; however, the normalised ratio between passive and active voice across the

whole corpus was 1:2.3, almost 21 times greater. The large difference in the ratio between clauses containing *world* and the whole of the FLOB shows that *world* has a strong aversion to being used in a passive context.

Figure 4.1 Voice distributions of *world* in sample corpus
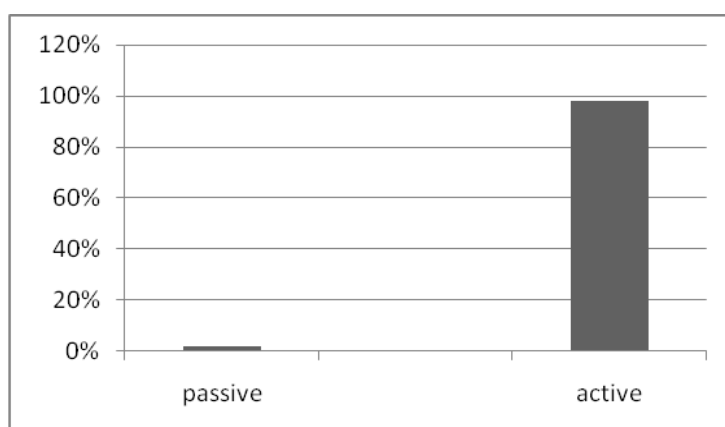
Of the eight instances of *world* used with passive voice, in four *world* was used as part of an adjunct—three times taking the form of *preposition + the + world* and once as postmodification. In the other four instances *world* was used as part of an object,—once as a pre-modifier, twice as a noun head, once as postmodification. Table 4.7 below shows a comparison of the distribution of passive voice vs. active voice in clauses containing *world* in the FLOB.

Table 4.7 A comparison of the distribution of Passive Voice vs. Active Voice in the verbal groups of clauses containing *world* and the FLOB

Voice Distributions of *world*		Distributions of Voice Across FLOB	
Passive	2%	Passive	31%
Active	98%	Active	69%

We next calculated the proportions of instances of *world* being used in clauses with perfect and non-perfect verbal groups, and then compared them with the ratio for the FLOB. Chart 4.2 shows that 10% (35/365) of instances of *world* were used with perfect aspect, while 90% (330/365) were used with non-perfect aspect, for a perfect-to-non-perfect ratio of 1:9. In the FLOB, 18% (7672/41747) of cases were used with perfect aspect, while 72% (34075/41747) were used with non-perfect, for a normalised perfect-to-non-perfect ratio of around 1:4.5, which is almost twice that of the distribution in clauses containing *world*. This finding shows that the distribution of aspect in verbal groups in clauses containing *world*, like the preference in the use of voice, does not correspond to the distribution in the corpus as a whole. *World* has a strong aversion to appearing in clauses that have a perfect aspect.

Figure 4.2 Aspect distributions of *world*

108
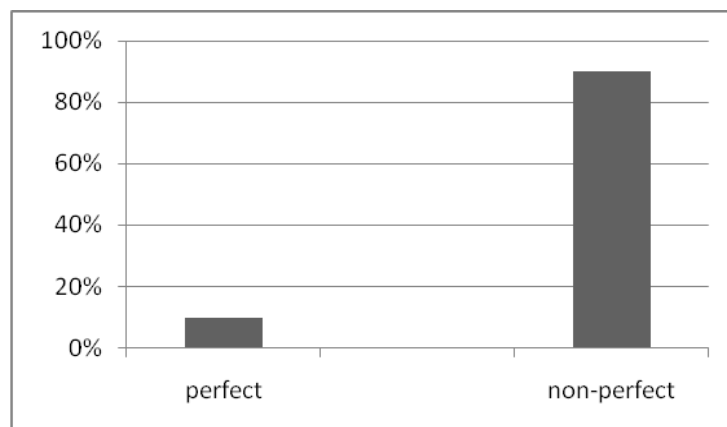
Based on these findings, the general colligational behaviour of *world* with regard to the tense, voice, and aspect of the verbal groups with which it occurs can be summed up as follows, after comparing with the distribution:

1. *world* does not have a significant tendency regarding the tenses with which it occurs. Its percentages of use with present, past and future tenses generally correspond to those for the FLOB;

2. *world* has a strong aversion to being used with passive voice, and a strong tendency to be used with active voice;

3. *world* has a strong aversion to being used with perfect aspect, and a strong preference to be used with non-perfect aspect.

4.3.2 Colligations of *world* in the clause

According to Hoey (2005:44), "A noun will always be part of some group or other word sequence and that group or word sequence will normally perform some function in a clause." It is thus reasonable to examine the distribution of a noun in terms of its occurrences within a clause or group. This section observes the colligational behaviour of *world* at the clausal level, and then its colligational behaviour at the level of group. Four major grammatical functions are considered in connection with *world*: as part of a subject, as part of an object, as part of a complement, and as part of a prepositional phrase functioning as an adjunct. Instance that do fit one of these four basic grammatical categories are simply analysed as Other. As long as the node word

appears in a piece of language functioning as subject, object, or complement, it is treated as belonging to that grammatical function.

Our use of these grammatical terms is generally in line with normal usage and the definitions in the Collins Grammar Dictionary, wherein subject refers to "the noun group that refers to the person or thing that is doing the action expressed by the verb" and object is defined as "a noun, noun phrase or pronoun that refers to a person or thing that is affected by the action of the verb (called the DIRECT OBJECT), or that the action is done to or for (called the INDIRECT OBJECT)". At this stage of the analysis, an instance of *world* was treated as subject whether it appeared as head of the nominal group (as Example 1 shows) or as part of an appositive subordinate clause to the subject (as Example 2 shows).

1) For Einstein, **the physical world** **[subject]** was an incarnation of reasonwhich, though manifest in various laws and principles, was inaccessible to the human mind in its profoundest depths.

2) **Graf, the girl who once ruled the world in awesome fashion** **[subject]** signalled her comeback by regaining her Wimbledon crown with a gutsy win over Argentinian star Gabriela Sabatini.

Complement, on the other hand, is defined as "an adjective group or noun group which comes after the verb and describes or identifies the subject." It normally follows the BE verb or other equative verbs, such as FEEL, BECOME, and SEEM.

Distinguishing adjuncts and postmodifying prepositional phrases when analysing data

is not as easy and clear-cut as one might think. Example 3 is clear-cut and unproblematic in its classification as prepositional object in a prepositional phrase that modifies its head noun—person.

3) You were my favourite person **in the world**.**[Postmodification]**

4) CGE is the largest water group **in the world**.

Example 4, while initially allocated to adjunct, raised interesting questions about the prepositional phrase, *in the world*. The most common use of this combination is to emphasise the nominal group proceeding to it—my favourite-in-the-world person—and it normally colligates such superlatives as *the best*, *the greatest* and *the largest*. However, one cannot simply classify *in the world* in a sentence such as *Vera Wang is the most famous designer in the world of fashion* into any category. At first glance, *in the world* in this case may seem to be a postmodification to the noun head *designer* in the nominal group that functions as complement. *In the world of fashion* can, however, be moved to the very beginning of the sentence to function as an adjunct, without greatly changing the meaning of the whole sentence. In this particular instance, *in the world* has an ambiguous analysis and both classifications make sense. There were several similar cases in our data; we grouped them into a third classification, "benign ambiguity,"[13] referring to instances that can be classified and analysed as both postmodification and adjunct, and excluded them when analysing the data.

We earlier compared features of tense with Halliday's normalised distributions and

---

[13]This term is suggested by my supervisor in a Skype meeting

self-calculated normalised distributions in the FLOB to investigate *world's* particular colligational preference. Similarly, this section compares the syntactic behaviours of *world* with those of other apparently similar nouns discussed in Hoey's work (2005: 46-49). In *Lexical Priming*, Hoey took four abstract nouns as comparison words. The grammatical distribution of these four words and the node word *consequence* are shown in Table 4.8 in comparison to *world* in the present study.

Table 4.8 A comparison of the grammatical distributions between of *world* in a clause and that of five words discussed in Hoey (2005)

	Part of Subject	Part of Object	Part of Complement	Part of Adjunct	Benign/others
world	27% (99/365)	33% (120/365)	19% (69/365)	14% (51/365)	7% (26/365)
consequence	24% (383/1615)	4% (62/1615)	24% (396/1615)	43% (701/1615)	5% (74/1615)
question	26% (79/300)	27% (26/300)	20% (60/300)	22% (66/300)	4% (13/300)
preference	21% (63/300)	38% (113/300)	7% (21/300)	30% (90/300)	4% (13/300)
aversion	23% (47/203)	38% (77/203)	8% (16/203)	22% (45/203)	8% (17/203)
use	22% (67/300)	34% (103/300)	6% (17/300)	36% (107/300)	2% (6/300)

Table 4.8 reveals there were no striking differences between *world* and the comparison words to *consequence*, except in the fourth grammatical function, adjunct, which shows a strong negative colligation between *world* and this function.

World and *consequence* showed a similar frequency of use as part of a subject—27% in our sample and 24% in Hoey's, respectively. However, while there was a clear negative colligation between *consequence* and the grammatical function of object, the distribution of *world* occurring in the case of object was almost the same as that of

112

other nouns in the table. With regard to the third grammatical function, complement, *world* had a tendency, though not as strongly as *consequence* or *question*, to serve in this function. Thus, the clause function colligations in which *world* was likely to appear can be summarised as follows:

1. There was a clear negative colligation between *world* and the grammatical function of adjunct. Only 14% of instances belonged to this function, almost three times less often than for *consequence*.

2. Regarding complement, while the tendency of *world* to occur in this function was not as strong as that which *consequence* exhibited, it still occurred much more frequently than did other parametric words, with the exception of *question*.

4.3.3 Colligations of *world* within the nominal group

This section examines the colligational features of *world* at the nominal group level, which include three grammatical possibilities: occurring as head of the nominal group, as a premodifier, or as part of the postmodification. For example:

5) They argue that the **world** has obligations under the Genocide
   Convention of 1948. [*world* **as head**]

6) The lad from a farm close to a one-horse stop called Peachester, near
   Brisbane, became a **world figure** with some of the best golf ever
   seen in the last round of a major championship.
   [*world* **as premodifier**]

7) Unfortunately he cannot pick Ian Baker-Finch, one of the best

known golfers **in the world**.

[*world* **as part of the postmodification**]

We analysed all nominal groups within the 365 instances from the FLOB containing the node word *world*. As before, its syntactic behaviour was compared with that of *consequence,* as analysed by Hoey (2005:49). The other four words used in Hoey's comparison were also used, to complement the result.

Table 4.9 A comparison of the grammatical distribution of *world* in the nominal group with that of *consequence*

	Head of nominal group	Premodifier of nomical group	Part of the postmodification of the nominal group
world	**33% (119/365)**	**25% (92/365)**	**42% (154/365)**
cosequence	**98% (1588/1615)**	**0.06% (1/1615)**	**2% (26/1615)**
question	92% (275/300)		8% (25/300)
preference	84% (253/300)	3% (8/300)	13% (39/300)
aversion	82% (167/203)	6% (12/203)	12% (24/203)
use	75% (226/300)	1% (2/300)	24% (72/300)

Interestingly, as Table 4.9 shows, *world* clearly differed in its grammatical distribution from *consequence*, despite their sharing a same colligational preference for the subject function. It differed strikingly from *consequence* and the other four nouns in two ways. First, *consequence* strongly tended (985) to occur as head of its nominal groups, as did the other words (75% to 92%), whereas *world's* tendency was only 33%. *World* occurred least frequently as part of the premodification among the three possibilities, it was overwhelmingly more likely to do so than the other five words, especially *consequence*, which showed almost no such tendency; 38% (34/92) of all

occurrences took the possessive construction (*world's*), while others took the noun modifier form (*world)*. The ratio between *world* as noun modifier and *world's* as possessive was approximately 1.7:1.

Second, while *preference*, *aversion*, and *use* occurred in postmodification between two times and twelve times more frequently than *consequence*, *world* showed an even stronger tendency, occurring 21 times more often than *consequence*. Within the 154 instances of *world* occurring as part of postmodification, 70% (108/154) occurred in the form of a prepositional phrase, dominantly represented by the prepositional phrase *in the world*, accounting for 34% of prepositional phrases (37 out of 108) and *on the world* (*stage*), accounting for 28% (30 out of 108).

The typical colligational behaviours of *world* at the nominal group level included:

1. a strong preference for being used as part of the postmodification of a nominal group;

2. a positive preference for appearing as part of the postmodification of a nominal group and proportionally frequently in the form, *in the world;*

3. a strong preference for being used as the premodification of a nominal group;

4. an unwillingness to occur in the possessive form (*world's*) when appearing as the premodification of a nominal group, but nearly twice the chance of occurring in the possessive form (*world's*) in a premodifier that of occurring as *world* as a noun modifier at this grammatical position; and,

5. a strong aversion to being used as the head of a nominal group.

So far, we have applied the "classic" approach as demonstrated in LPT to examine the

colligational behaviour of *world*. As discussed in Chapter 2, a co-colligational approach is also used in the study to detect the relation between inter-colligation and intra-colligation. The former refers to colligational investigation at the nominal group level; the latter refers to colligational investigation at the grammatical function level and clausal level. Now that we have examined the colligational preferences and aversions of *world* with regard to its grammatical functions and grammatical positions in a nominal group, it is possible to look at how these colligations interact with its distribution within subject, object and complement. We will then look at the two dominant adjunct forms which were made up of two typical prepositional phrases.

4.3.4 Grammatical distribution of *world* in the nominal group within subject

Table 4.10 Grammatical distribution of *world* in the nominal group within subject

Subject (99)	postmodification (46/99)	prepositional phrase: 50% (23/46)
		a world of: 50% (23/46)
	premodification (17/99)	world as modifier: 51% (9/17)
		world's: 49% (8/17)
	head (32/99)	definiteness: 90% (29/32)
		indefiniteness: 10% (3/32)

The first thing to note about *world* in the nominal group within subject is its positive colligation for functioning as part of postmodification. Of the 99 instances in this category, 46 instances occurred in this grammatical position (46%). The dominant postmodification patterns were displayed by prepositional phrase *in the world* and the structure of *of + world*. The distribution of these two patterns was even (1:1 ratio).

The second most frequent function with which *world* occurred as part of a subject was that of functioning as part of a head. Ninety percent of these cases occurred in a pattern with definiteness, e.g., with *the* or *my* preceding it, while 10% occurred in a pattern with indefiniteness, e.g., with *a* and *one* preceding it. When *world* was part of subject, premodification occurred in only 15% of all instances.

4.3.5 Grammatical distribution of *world* in the nominal group within object

Although Section 4.3.2 showed that the use of *world* as part of an object was not in itself colligationally interesting, such instances were common. The word had neither a preference nor an aversion for the grammatical function of object, when compared with *consequence* and the other four abstract nouns in Hoey's work; however, this colligational feature of *world* was its most frequent grammatical function in our data, its uses within object deserve closer investigation to see whether there were other, more distinctive colligational features associated with them.

Table 4.11 Grammatical distribution of *world* in the nominal group within object

Object (120)	postmodification (41/120)	Prepositional phrase: 58% (23/41) …of: 42% (18/41)
	premodification (38/120)	*world* as modifier: 55% (21/38) *world's*: 45% (17/38)
	head (41/120)	definiteness: 80% (32/41) indefiniteness: 20% (9/41)

Table 4.11 shows there is a 1:1 ratio between of instances of *world* functioning as part

of postmodification and as head. On the other hand, only 38 of 120 instances (32%) occurred within premodification. Among these 38 instances, 45% (17 of 38) occurred in the possessive form *world's* and 55% (21 of 38) occurred when *world* served as a modifier. Patterns associated with definiteness in head were four times more frequent than patterns associated with indefiniteness.

4.3.6 Grammatical distribution of *world* in the nominal group within complement

When *world* was head of a nominal group serving as complement, it appeared to colligate with non-specific deictics more often than it served other grammatical functions, as can be seen in Table 4.12. The ratio between definiteness and indefiniteness was 1:3, as opposed to 9:1 and 4:1 when appearing as head of the nominal group functioning as subject and object.

Table 4.12 Grammatical distribution of *world* in the nominal group within complement

Complement (69)	postmodification (26/69)	Prepositional phrase: 89% (23/26) …of: 11% (3/26)
	premodification (23/69)	*world* as modifier: 25% (6/23) *world's* : 75% (17/23)
	head (20/69)	definiteness: 25% (5/20) indefiniteness: 75% (15/20)

When appearing as postmodification of the nominal group in complement, *world* seemed to have a stronger preference for occurring in the prepositional group form

than in the other two grammatical functions. Similarly, in connection with its premodifying function, the ratio between premodification realised by *world* itself and the possessive *world's* was 3:1, as opposed to 1:1 when functioning as part of subject or object.

4.4 A summary of *world*'s colligational behaviour

In the above three sections we examined the general collocational and colligational features of *world*. With regard to its general collocational profile, we investigated collocates occurring on both the left and right sides of the node word *world*. The collocates at the L2 position were dominantly occupied by prepositions, while collocates at the R1 position were dominantly by superlatives. Since these features were more likely to be colligational, we then investigated the colligational behaviour of *world* in nominal groups when serving as part of different grammatical functions, namely subject, object and complement; as mentioned before, the adjunct function will be discussed in the next chapter. One striking finding was that *world* occurred as part of postmodification in the nominal group almost 21 times more often than did *consequence* and twice as often did *use* in Hoey's work.

Likewise, there was also a clear positive colligation between *world* and the grammatical position of premodification. The other nouns occurred in this grammatical position with lower frequency, ranging from 0.06% to 6%, with *question* occurring not at all at this grammatical position. *World*, on the other hand, occurred within premodification roughly a quarter of the time.

However, when we studied how grammatical position with the nominal group interacted with the grammatical functions, a significant difference was found in the tendency of *world* to be used as premodifier. In only 15% of instances was *world* used as part of premodification in a nominal group serving as subject, almost half the rate when it occurred in object and complement. We then moved on to detect close pattern features of *world* in the nominal group in connection with different grammatical functions, and found no significant differences between the two patterns of postmodification—i.e., prepositional patterns such as *in the world* and *… of world* patterns—found in nominal groups of subject and object functions. Similarly, there were no significant differences between the two patterns of premodification found in nominal groups of these two functions. However, the comparison between patterns was significant in complement function. The ratio between prepositional and *…of world* patterns in postmodification was around 9:1, while the ratio between *world* as premodifier and *world's* as premodifier was around 3:1. However, the ratio of the two patterns of postmodification was 1:1 in subject and 1.5:1 in object, while the ratio of the two patterns of premodification was 1:1 in subject and 1.2:1 in object.

These preferences and aversions are summarised below:

1. when appearing as part of postmodification, *world* has a strong preference for being used within a subject;
2. when appearing as part of premodification, *world* has a strong preference for being used within a complement;
3. when appearing as head, *world* has a strong preference for occurrence within a subject;

4. when appearing as part of postmodification, *world* is likely to occur in the form of a prepositional phrase; and,

5. *world* is likely to associate with definiteness when appearing as part of head within a subject or an object, and with indefiniteness when appearing as head within a complement.

4.5 A discussion of data

We saw in Section 4.2 that a group of prepositions (e.g., *around*, *in*, *over*, and *of*[14]) occurred dominantly at the L2 position of *world*, and a group of superlatives (e.g., *largest*, *most* and *biggest*) occupied largely the L3 and R2 positions. These occurrences imply not only a collocational behaviour but also a colligational feature, even if some words were not strictly categorised as embracing grammatical functions. The findings support our long discussion on the relation between lexis and grammar, in Chapter 2, that there is no 'in the first place' and 'in the second place' order in the process of language uses; i.e., human beings do not first choose a word to fill the slot determined by grammar and no set grammar fits every word in every context.

It is widely recognised by Sinclairian linguists that lexis and grammar are correlated, and that there is a need to account for the existence of this correlation and how it works. Lexical priming is intended to serve as a bridge between lexis and grammar. If priming operates with regard to collocation, then it must involve grammatical constraints. In our data, *world* was primed to occur with *the* at L1 and the sequence

---

[14] As noted previously that we agree with Sinclair's claim concerning the grammatical category of *of* that it is not a true preposition; here, however, we were including it as a preposition because when it collocates with *world* to form the combination *of the world*, it is normally serving a prepositional function.

the world was primed to occur frequently with *in* at L2; thus, the sequence *in the world* was primed to occur with superlatives such as *largest*. Following this line of thought, the basic idea of colligation is formed in that a lexical item may be primed to co-occur with another lexical item. This lexical item may also be primed to occur or to avoid occurring in or with a grammatical function.

The examination of *world*'s colligational behaviour reinforces this claim. We found *world* had a significant tendency to occur in clauses in the present tense (sig.=0.000) and a strong aversion to being used with passive voice and with perfect aspect.

Our study of 365 examples of *world* occurring at the clausal level revealed that *world* did not have a clear preference for being used within subject and object compared with Hoey's five abstract nouns, even if the majority of instances containing *world* in our sample occurred within an object. The node word *world* neither avoided nor favoured these two grammatical functions, which turned out to be a unique colligational feature of the word. Though *world* did not exhibit a preference or aversion for the functions of subject and object, it showed a clear negative colligation with the grammatical function of adjunct and a relatively stronger positive colligation for occurring within a complement.

Our examination of instances containing *world* at the group level indicated it had a preference for different positions within different grammatical functions: occurring as part of a postmodificaiton within a subject, as part of a premodification within a complement, and as head within a subject. *World* also had a preference for the pattern it appeared in when occurring at different grammatical position. For example, *world*

preferred to appear in the phrase *in the world* as part of a postmodification. In this sense, colligation was taken to mean a relation holding between a word and a grammatical pattern (Halliday 1959). This is an important extension that interprets colligation as going beyond traditional grammatical relationships. There is no single grammar to a language, but a "multiplicity of overlapping grammars that are the product of the attempt to generalise out of collocations" (Hoey 2005: 47). This does not mean that grammar is devalued; however, primed collocations are acquired through others' utterances, and those source utterances might be tempered by the grammars the language users have encountered.

In this chapter we have witnessed how collocation and colligation interconnect. We have also noted that a word's patterns of use are characteristically controlled by its colligations. In the next chapter, we shall explore the interrelation between *world*'s collocation, semantic association and colligation. Complex issues concerning collocational behaviour within different colligations are addressed in the following chapter.

CHAPTER 5

World in Complex Situations

5.1 Introduction

To detail the collocational behaviour of *world*, this chapter examines it in combination with its colligational profile. It investigates the collocates of *world* within different grammatical functions, appearing at different grammatical positions. These collocates are then grouped into semantic associations, defined in LPT as "a lexical choice drawn from a recognizable semantic set or field" (Hoey 2014:182). Furthermore, textual semantic association will be studied by looking at neighbouring words further from the node word. The identified semantic associational behaviours of *world* will be compared with the collocational, colligational and semantic associational behaviours of its Chinese equivalent 世界 *shi4jie4* (world), to answer the second research question: "In what ways do English and Chinese differ or show similarity with regard to the characteristic primings for lexical and syntactical behaviour?"

5.2 Semantic associational profile of *world*

This section examines the general profile of *world* from three perspectives. First, it groups collocates of *world* at respective positions into semantic associations. Second, it detects how the two most frequently prepositional phrases containing *world* were associated with different meanings. The collocates co-occurring with the phrases will be formed into semantic sets by associating them with different meanings. Last, the

possessive form of *world* (*world's*) will be studied with regard to the semantic associations revealed by its most frequent collocates.

5.2.1 A general semantic associational profile of *world*

As shown Table 4.1, the top collocates occurring at the L1 position of *world* were *Third*, *Second*, *First*, *real, outside, Arab, whole, the, modern, new, third, Western, natural, developing, second,* and so on. We can group a set of words, including *Third*, *Second*, *First*, *third* and *second* into a semantic association (and colligation) of Ordinal Number and the words *Western* and *Arab* into a semantic association of AREAS/COUNTRIES. The L1 collocates listed above were frequently-used candidates, with LogDice scores at or over 6.5. To investigate whether collocates with lower collocational strength were also members of *world*'s semantic association at the L1 position, we looked at L1 collocates with a LogDice score below 4.5[15]. We found collocates such as *Greek* (4.291) and *England's* (4.083) were also associated with the semantic meaning of AREAS/COUNTRIES. Collocates such as *today's* (4.402), *contemporary* (4.459) and *everyday* (4.161), together with the frequent collocate *modern* seemed to form a semantic association of POSITIONING IN TIME.

There will always be co-occurrences that cannot be accounted for in terms of collocation, such as rare co-occurrences at the L1 position. Collocation is—no matter what kind of definition it is given by researchers—a statistical phenomenon

---

[15] We noted in Chapter 3 that a two-score difference of LogDice score could manifest a strong significance in terms of collocational strength.

manifested between words. Statistically non-significant co-occurrences were not considered while discussing collocation. Semantic association, thus, is a "necessary generalisation" between words and semantics. It is probable (but not theoretically necessary) that collocations are primed first and that semantic commonality between collocates produces an abstract priming, whether due to self-reflection or to encountering co-occurrences sharing the same semantic features. It is also possible that the semantic associations of a word or word sequence are primed first, with the collocates being primed by an individual's selection from the semantic set. With regard to the first possibility, Sinclair (1999, 2004a) used the term semantic preference to describe a semantic tendency in which a word is likely to appear. With regard to the second possibility, Sinclair (ibid) used the term semantic prosody to account for individuals' motivation for selecting words in a particular communicational context. We have discussed, in the literature review, that Sinclair and Hoey understood semantic association/preference in the same way, with the latter being applied only to get alleviate the occasional confusion arising from psychological preference. However, Hoey seemed to bridge semantic association and semantic prosody with the characteristic notion of priming, which moves outwards from specific words to specific semantic sets to permit creative word choices—e.g., *century by century* or *millennium by millennium*—in Taylor's (2008) N by N structure

Table 4.2 in Chapter 4 lists the ten most frequent collocates at the L2 position. The majority of these collocates were prepositions, including *around*, *throughout*, *outside*, *over*, *in*, *of*, *round*, *across*, and *into*. These collocates, though better categorised as grammatical words, could also be grouped into a semantic association

of POSITIONING IN LOCATION, represented by prepositions such as *around*, *outside*, and *in*. The prepositions *during* and *after*—not shown in Table 4.2, but still a frequent collocate at the L2 position of *world* with a LogDice score of 5.074—seemed to be associated with a semantic meaning of TIME PERIOD.

As has been discussed in the previous chapter, a word is primed to co-occur with another word (collocation), and subsequently this word sequence is primed to collocate with other words. This feature is described in LPT as *nesting*, "where the product of a priming becomes itself primed in ways that do not apply to the words making up the combination" (Hoey 2005: 8). The highly frequent uses of prepositions at the L2 position need to be accounted for with the feature of nesting to draw a comprehensive picture of *world*'s semantic association at the L2 position. We note that if we take nesting into consideration at this stage, prepositions such as *in* cannot be grouped to the semantic association of POSITIONING IN LOCATION, but to the semantic association of POSITIONING IN TIME, as in the nesting of *in the world war*. At the current stage we restricted our discovery to the L2 position without taking nesting into account and will come back to this issue later in this chapter.

We noted in Chapter 4 that collocates at the right side of the node word *world* were not as various as those on its left side. Nevertheless, some collocates can also be grouped, where appropriate, as members of particular semantic associations. At R1 positions, the collocates with higher LogDice scores were *war, Cup, Bank, Champion, war, largest, trade, countries, Health, economy, Broadcast, title, market, champions*, and *championship*. These collocates showed a semantic association that

might be labelled as COMPETITION, represented by words of *cup*, *Champion*, *champions* and *championship*; ECONOMY, represented by *Bank*, *trade, economy*, and *market*; EVALUATION, dominantly represented by the superlative *largest* and supplemented by members with lower LogDice scores, such as *biggest* (6.324), *greatest* (5.776) and *oldest* (4.588). As with the semantic association grouped at the L1 position, there were also less common collocates occurring at R1 position that were associated with the same broad semantic meaning as that of highly frequent collocates. Semantic association serves as a bridge to explain the existence of less frequent collocates or indeed non-collocates to the node word. We can use this finding as further evidence in support of the claim of priming in Hoey's theory: "that priming moves out from collocations to semantic association" (Hoey 2005:18). For instance, in our data, *largest* was a highly frequently-used R1 collocate to *world*, and also a member of the semantic set forming a semantic meaning of EVALUATION. Thus, *world's largest* in turn became a strong collocation. According to the LPT, primings "move outwards from specific words to the semantic set, and in so doing permit creative choices to be made that in themselves reinforce the more general priming" (ibid). Following this line of thought, we could use priming to explain the existence of *oldest*, a less frequent collocate but a member belonging to the same semantic association of EVALUATION, by saying that people encounter *world + largest* (or *biggest* and *best*) very frequently to form a solid knowledge for this semi-fixed phrase, and when they want to use the same format to evaluate something *old*, they "are primed by" the semantic meaning formed by *world + largest* and then make a creative choice to replace *largest/biggest* or *best* with a much less frequent item (e.g., *world's furriest*). Superlatives such as *largest*, *biggest* and *best* strictly belong to colligational features. In this case, the distinction between colligation and

collocation was not clear-cut, indicating that grammar is sometimes lexicalised and cannot be studied separately from lexis.

In examining (in Chapter 6) *世界 shi4ji4,* we will look at whether *世界 shi4ji4* (world) had semantic associations and, if so, whether these semantic associations occur with colligations and there is nesting.

5.2.2 A semantic associational profile of prepositional phrase containing *world*

We noted in Chapter 4 that *world* occurred dominantly in prepositional phrases that, in turn, had their own collocational, colligational semantic associational features. This is known as *nesting*, where a combination of words may have its own collocations, colligations, semantic associations, and so on.

In the BNC, prepositional phrases containing *world* (25,234) accounted for 43% of all concordance lines containing *world* (58,496). Of these instances, *in the world* and *of the world* occur dominantly. There were 7,236 instances of *in the world* and 6,409 instances of *of the world* occurring, accounting for about 29% and 25% respectively of the total prepositional combinations containing *world*. Altogether, the two nestings made up more than half of the prepositional instances of *world*. They deserve closer attention because they may indicate other, more distinctive collocations, colligations and semantic associations with them.

5.2.2.1 A semantic profile of *in the world*

We examined collocates of *in the world* within a five-word span; the distributions are displayed in Table 5.1 below:

Table 5.1 Collocates of *in the world* within a five-word span

Rank	Candidates	Cooccurrence Count	Score
1	anywhere	230	9.588
2	largest	222	9.287
3	Cup	159	8.622
4	biggest	122	8.581
5	best	453	8.568
6	greatest	102	8.269
7	finest	67	8.171
8	richest	52	8.05
9	oldest	57	7.985
10	most	672	7.876

Table 5.1 shows that all but two of the collocates of *in the world* formed a semantic association of SUPERLATIVE (and also a colligation). The tendency of *in the world* to occur with SUPERLATIVE was far stronger than that of the single node word *world*. This finding conforms to Hoey's claim that primings "nest and combine" (2005: 10-11). The product of nested and combined word sequences were earlier defined as nesting in Hoey's theory, which argues that the lexical behaviour of nesting does not apply to the words constituting it. Hoey used the example of *in winter* to illustrate this claim, in that *winter* collocates with *in* to produce the phrase *in winter*; however, this phrase has its own colligations with a number of forms of the word *BE*, which are separate from those of either *in* and *winter*. In addition, *in winter* is likely to be primed (for *Guardian* readers) to occur in clauses with the present

tense, while a similar nesting of *in the winter* is likely to be primed to occur in clauses with the Past tense.

Hoey's definition of nesting works for the instance of *in winter* in his discussion but does not strictly work for our example of *in the world*. The question lies in how we interpret the *distinctive features* between nesting and its components. In Hoey's example, *in* and *winter* do not share any collocational and colligational feature with *in winter* as a combination. However, in our example, *world* and *in the world* do share the same lexical behaviour. In respect of collocation, they share the same collocates of *largest*, *biggest* and so on; in respect of colligation, they share the same semantic association (and also colligation) of superlatives. Where the component *world* and the word sequence *in the world* differ is only in their collocational strength of co-occurring with their shared collocates.

Based on the claim proposed above, we offer the following tentative modification to the concept of semantic association in LPT—there is a possibility that a nesting may share some of its collocations or semantic associations with its components; where they may differ, however, is in the strength of their tendency to collocate or form a semantic associations; in the case of collocation, they may differ greatly in collocational strength (which could be revealed with LogDice score). For example, the LogDice score of *the world* collocating with *largest* is 7.633, while the score increases to 9.287 when *world* collocates with *largest* in the nesting of *in the world*. This phenomenon occurs with every component of the semantic set forming the semantic association of SUPERLATIVE.

5.2.2.2 A semantic profile of *of the world*

Examination of the collocates of the nesting of *of the world* gives more evidence to support our proposal. Table 5.2 below seems to confirm that, though sharing the same semantic association of SUPERLATIVE with the component *world*, *of the world* had a stronger tendency (7.087) to form this kind of semantic association, albeit less strongly than that in the case of *in the world*.

Table 5.2 Collocates of *of the world* within a five-word span

Rank	Candidates	Co-occurrences	Score
1	parts	490	9.853
2	rest	495	9.558
3	News	135	8.895
4	largest	91	7.973
5	Cup	55	7.793
6	nations	321	7.751
7	part	94	7.537
8	population	147	7.399
9	view	53	7.314
10	greatest	78	7.295

Our finding does not contradict the claim in LPT that "the product of a priming [nesting] becomes itself primed in ways that do not apply to the individual words making up the combination" (Hoey 2005:8) because priming in this theory includes various parameters: collocations, colligations, and semantic associations. What we note here is that even in cases where the node word and its nesting shared the same semantic association, and even where the members of the semantic sets forming this

semantic association were similar, language users were still primed differently for the node word and its nesting in terms of strength and tendency.

In Chapter 6, we will look at 世界 *shi4jie4* (world) to see whether it also has nesting and, if so, whether its collocations, colligations and semantic associations were different from or stronger/weaker than the collocations, colligations and semantic associations of the word on its own.

5.2.2.3 A semantic associational profile of *world's*

We saw in Chapter 4 that *world* was dominantly primed for occurring as part of a postmodification and as part of premodification. We also concluded that the former function was mainly realised through prepositional phrases, while the latter function was evenly realised through *world* as an individual modifier and *world* occurring in a possessive form (*world's*). In above sections, we have outlined a general lexical picture of two prepositional phrases *in the world* and *of the world*, regardless of its positions in a clause (we will detail the lexical behaviour of these two phrases at different grammatical position when occurring within different grammatical function in Section 5.3). In subsequent sections, we will carry out detailed studies of the collocational and semantic associational features of *world's and* address other situations where *world* occurs as part of premodification in Section 5.3. Results are displayed below in descending order of Log-Dice score:

Table 5.3 Collocates of *world's* within a five-word span

Rank	*World's*	Occurrences	Score
1	population	93	9.75
2	market	59	9.11
3	company	57	8.89
4	country	46	8.87
5	producer	40	8.69
6	people	33	8.30
7	nation	28	8.21
8	press	28	8.17
9	problem	37	8.16
10	player	28	8.10

Collocates appearing immediately after *world's* (R1) generally fell into three semantic associations: Human Beings (represented by words such as *population*, *producer*, *people* and *player*), Financial Organisation (represented by words such as *market* and *company*), and Country (represented by words such as *country* and *nation*).

The word *problem* did not belong to any of the above-mentioned sets. However, due to its relatively higher LogDice score, we took a closer look at the 37 instances of this kind in case there were other semantic features associated with them. Excluding seven instances in which *world's problem* functions as part of the a postmodification of an Object or Subject rather than a noun head, we have 30 valid instances left to examine. Of these instances, 20 functioned as an Object in a clause and collocated

with verbs forming a semantic association of SOLVING, represented by *solve* (11 hits) and verbs such as *tackle* and *sort out* that all together formed a semantic association of TAKING ACTION. There were seven instances of *world's problem* appearing for the function of Complement, collocating with words stating the practical problems occurring in the world, such as *overpopulation*, *pollution*, *environment*, *economy* and so forth.

The remaining three instances functioned as a Subject followed by the co-occurrences of *awful*, *big* and *insoluble*, forming a semantic association of NEGATIVE EVALUATION.

In our data there were few shared collocates between grammatically related forms of *world* and *world's*. With the sole exception of *market*, there were no frequently-used collocates that were shared by the two forms containing *world*. This finding conforms with Jantunen's (2017) extension of priming—morphological primings. Jantunen showed, in a case study of a Finnish lexical item *kello* ('what, time, o'clock), that morphological priming (inflectional priming to be specific) plays an important role in the learner language phraseology.

Chinese has a morpheme 的 *de* similar to the morpheme *'s* in English, in that it serves the same grammatical function (possessive). As we have for *world's*, we will also investigate the lexical and grammatical features of 世界的 *shi4jie4de* to detect whether this kind of morphological priming also occurs in Mandarin Chinese.

5.3 Collocational profile of *world* when occurring within Object

Chapter 4 indicated there was no clear preference or aversion for *world* occurring within an Object, compared with other nouns studied by Hoey (see Table 4.8); however, as mentioned previously, this is itself a colligational feature. In this section, we examine in detail the collocational and semantic associational features of *world* within an Object function in terms of three grammatical functions—premodification, head, and a postmodification.

5.3.1 Collocations and semantic associations when occurring as noun head

Sketch Engine shows the verbs most frequently preceding the node word *world* when it functioned as a noun head in a nominal group within an Object. This is displayed in Table 5.4 below, with numbers of occurrences and LogDice scores provided in separate columns.

Table 5.4 Top frequent verbs collocating with *world* as head of Object

Rank	verbs with *world* as object	Occurrences	Score
1	enter	118	8.07
2	change	167	8.06
3	travel	56	7.83
4	inhabit	55	7.74
5	create	131	7.61
6	save	61	7.40
7	dominate	74	7.31
8	explore	42	7.26
9	divide	45	7.23
10	rule	33	7.21
11	shock	30	7.19
12	perceive	32	7.00
13	shape	28	6.68
14	think	89	6.83
15	understand	44	6.79
16	transform	26	6.75
17	view	27	6.71
18	watch	44	6.64
19	imagine	26	6.62
20	conquer	20	6.62
21	love	30	6.59
22	face	44	6.55

These verbs can be grouped into four different semantic associations: SHAPE (represented by words such as *divide*, *shape,* and *transform)*; GOVERN (represented by words such as *dominate* and *rule*); to KNOW (represented by words such as *perceive* and *understand*); and ACTIONS TAKEN ON THE WORLD (represented by words such as *explore* and *conquer*).

One of one's basic choices when using a noun head in a nominal group is between definiteness and indefiniteness. The ratio between definiteness and indefiniteness when *world* appeared as noun head within an Object was 4:1 in our data. The dominant indefiniteness that preceded *world* was *a*. The combination of *a + world* occurred as part of two grammatical patterns: one followed by verbs that functioned

as a   postmodification of *world*, for example *a world dominated by…*, and the other pattern with *of* + N subsequent to it, such as *enter a world of fantasy*.

5.3.2 Collocations and semantic associations when occurring as part of postmodification

Another grammatical position as highly distributed as noun head within an Object was postmodification, which was dominated by the prepositional phrase *in the world*, and the pattern of head noun + *of the world*, such as a *tragedy of the world*. When existing in the form of head noun + *of the world*, the preceding words tended to be associated with a meaning of PART, such as *rest* and *parts*.

5.3.3 Collocations and semantic associations when occurring as premodification

World had a clear aversion for occurring as premodification within an Object. Of the 38 instances found in the sample corpus (made up of 365 instances containing *world*), 21 instances were found to occur as individual words, i.e., *world* + noun head, such as *world community,* while 17 instances were found to occur in the form of *world's*, such as *world's population*.

5.4 Collocational profile of *world* when occurring within Subject

While *world* did not show a distinct preference or aversion for appearing within a Subject compared with nouns examined in Hoey's work, it showed a clear

preference/avoidance for particular grammatical positions within the Subject function. These preferences and avoidances are detailed below.

5.4.1 Collocations and semantic associations when occurring as noun head

As mentioned above, one of the basic choices we make whenever we use a nominal group, especially when choosing a head noun, is between definiteness and indefiniteness. A natural first step in examining *world* as a Subject was therefore to consider the patterns of definiteness associated with it. The behaviour of *world* in comparison with words studied by Hoey was also considered, lest we fail to draw a correct conclusion. As a head of a nominal group within a Subject, *world* also appeared to have a positive colligation with definiteness more often than other nouns. The ratio between definite and indefinite Subjects with *world* was 9:1, as opposed to 3:1 for *consequence*, 4:1 for *preference*, and 2:1 for *use* (comparison data extracted from Hoey 2005:56). Another colligation can be detected by looking more closely at how definiteness was realised in the Subject nominal groups. Our sample revealed three main ways in which a nominal group may be made definite: with the definite article (*the*), with a possessive pronoun (e.g., *his*, *her*, *our*), and with a determiner (e.g., *this*, *that*). In comparison with Hoey's five abstract words (*consequence*, *question*, *preference*, *aversion* and *use*), *world* overwhelmingly favoured occurring with the two determiners, *this* and *that*. In addition, *world* also had a strong preference for occurring within a possessive structure. However, occurrences of *world* favouring *the* as a marker of definiteness were less frequent than those of *consequence*, *question* and *use* (see Table 5.5)

Table 5.5 A comparison of *world* and five other nouns in respect of definiteness

	the	possessive	this, that
world	52%	31%	17%
consequence	99%	1%	
question	96%	4%	
preference	25%	70%	5%
aversion	27%	73%	
use	64%	36%	

A noun head is the centre of a nominal group. When the nominal group appears within a Subject, it becomes the descriptor of the clause. The semantic association summed above implies that *world* was primed by English speakers as an entity that was likely to change or to be changed when the speakers intend to use *world* as part of a Subject.

Compared with verbs appearing at the left position to the node word *world*, the verbs appearing to the right of *world* seemed not to be grouped into semantic sets (see Table 5.6 below), with the possible exceptions of *change*, *become*, *turn*, *shrink*, and *begin*, which could be grouped as forming a semantic association of CHANGE.

Table 5.6 Top frequent verbs collocating with *world* when Subject

verbs with *world* as subject	Occurrences	Score
change	61	7.41
know	84	7.15
seem	98	7.03
see	71	6.85
go	124	6.34
become	61	6.04
fall	33	6.03
turn	32	6.02
shrink	12	5.96
wait	18	5.95
need	34	5.94
exist	19	5.91
have / has / had	777	5.88
begin	41	5.83
owe	12	5.78
be	2301	5.76
lose	19	5.75
come	79	5.65
look	39	5.63
want	27	5.61
do	123	5.56
depend	14	5.51

5.4.2 Collocations and semantic associations when occurring as postmodification

Instances containing *world* appearing as part of a postmodification within a Subject had a strong preference for occurring as the prepositional phrase *in the world*. The noun heads that were in this way postmodified tended to be associated with a meaning of ORGANISATION, dominantly occupied by *church* and *university*. In addition, noun heads postmodified by *in the world* had a positive preference for being modified with superlatives, as in the example of *the oldest university in the*

world… Instances containing this structure were found to occur in 83% of instances (19 of 23).

With respect to the second dominant form of postmodification—*of the world*—a semantic associational behaviour can be found, in that the nouns postmodified by this phrase were associated with a meaning of ECONOMY, represented by collocates of *economy* and *bank*.

The semantic associational behaviour detected from *in the world* and *of the world* showed that within the same colligation, the semantic association differed strikingly due to a component change to a nesting. The difference evidenced in the corpora reveals that within each context or semantic environment, individuals were primed with distinctive collocations.

5.4.3 Collocations and semantic associations when occurring as premodification

The distributions of *world* and *world's* were even (51% vs. 49%) when serving as a premodification in the Subject. Word types modified by *world* were sparse and did not permit grouping into any semantic set forming a particular semantic association. On the contrary, words modified by the possessive form *world's* tended to form a semantic association with COMMUNAL PROBLEM, illustrated by *pollution* and *unemployment*.

5.4.4 Complex issues concerning semantic associations

Above, we examined the collocational, colligational and semantic associational features of *world* occurring within a Subject by looking at instances that occurred statistically more often than others. In this section, however, we look at combinations with few data to investigate whether these combinations exhibited particular behaviours and whether the components shared collocates and semantic associations with the combinations that made them. Furthermore, we also investigate the collocational and semantic associational behaviours of *world* beyond its close neighbours (the less-close context) to see whether the word was affected by its less-close neighbours as well.

5.4.4.1 Complex issues concerning combination

world + **SHIRNK**

We began by looking at the combination of *world* + SHRINK—chosen (along with *world* + OWE in later paragraphs) for its less frequent occurrences in the BNC—to detect whether typical semantic features were occurring. The capitalised SHRINK was refers to every kind of word form of *shrink*, including *shrinks*, *shrank*, *shrunk*, and *shrinking*. While different verb forms will have different collocates and are likely to reveal different semantic associations, we here took all verb forms of *shrink* because some associations and collocations of the components of this combination are shared, based on their LogDice score, and because the number of instances

containing *world* + SHRINK was too small for there to be strong statistical differences when determining the collocational and semantic associational features.

We found 12 instances of *world* + SHRINK in the BNC. After excluding one instance in which *world* was used as part of the postmodification of a noun phrase functioning as a Subject (see example 1 below), we had 11 instances left to be examined.

1) Britain's trade deficit with the rest of the world shrank dramatically as output from factories rose.

Of the remaining 11 instances, three expressed a meaning of *world*—the planet we were living on—while the rest expressed another meaning of *world* referring to a way of life one had or is used to having, or to a particular group.

2) Her **world** had shrunk to the few things which still had interest for her. [The way of living that she had been used to]

3) 88open's **world** shrinks painfully with every defection from the Motorola Inc 88000 RISC camp, or the adoption of a dual-architecture strategy by a formerly loyal vendor.

[88open refers to an industry standards group (the 88open Consortium) set up by Motorola in 1988 to standardize Unix systems on their Motorola 88000RISC processor systems.]

In two of the three cases where *world* was used to refer to everything living on the planet, the word *technology* occurred within a five-word span. Taking the further environment (e.g., a 15-word span) into consideration, *world* in these two sentences also occurred with words such as *telephone* or *global system*, which could be used to

generalise a semantic association of HIGH TECHNOLOGY. Studying a word's collocational and semantic behaviours within a further span has been done by Hoey (2014), who noted three important observations:

> The first important observation…was that each collocation holds only for the grammatical form quoted;
> The second important observation was that the neighbours of a word may strongly affect the sense of the word in question;
> The third important observation… [was that] not only do a word's different neighbours affect the word's likely sense but the different ways in which the same neighbor (neighbour) was used may do the same.
> (ibid:181-182)

Hoey claimed that a word's relationship with its neighbours "was not exhausted by this narrow span" (ibid:182). To examine why the mis-selected choice of *crude forgeries* (where *crude forges* shall be the right one)[16] was not detected by people who were shown the passage, Hoey sketched out some of the ways in which a word's sense might be affected by its neighbours. He found out that the less immediate environment, which was the combination of FUEL semantic association, was liable to make the reader take the phrase *smoke from the crude forgeries* as *smoke from the forges*. Since the difference between forges and forgeries lies in the ending of *eries*, Hoey examined the ways in which *–eries* was used in English by investigating 118,932 tokens of words ending with *–eries* in combined the corpora of the *Guardian* and the BNC. The data showed that people were characteristically primed to associate the ending of *forgeries* with a meaning CRIME, SIN, UNPLEASANT, and PRODUCTION. The words forming these semantic associations occurred 11 or more places from the node word *forgeries*; the semantic association was also evident in

---

[16] In the village of Chilling there were more lessons, this time in metalwork. From out of this tiny mud-hut hamlet comes the most beautiful beaten bronze, copper and silver, found cladding traditional kitchen stoves across Ladakh. Smoke from the crude forgeries rose over the village as I picked my way carefully down the mountain between twisted trunks of willow trees. (The Independent on Sunday, 2 November 2008, p.69 cited in Hoey 2014)

many cases of *forge*.

To Hoey, semantic association can be studied within a further span; semantic features are not exhausted at a four- or five-word span. This is defined as textual semantic association in the LPT: "Words (or nested combinations) may be primed to occur (or to avoid occurring) in specific types of semantic relation, e.g., contrast, time sequence, exemplification, comparison, cause-effect, problem-solution" (Hoey 2005:115).

Textual semantic association is an extension of the notion of semantic association and in some of its manifestations might be distinguishable from the kinds of semantic association looked at earlier in this chapter. The claim holds that semantic relations or discourse patterns may be textual. Relationships may be reflected in clauses or parts of clauses (or larger chunks of text), or may reflect the interaction between the writer and the reader.

Returning to the remaining nine instances containing *world* + SHRINK, five instances had co-occurrences with possessive pronouns (among which three instances occurred with *her*, one with *my,* and one with *88 open's*), thus forming a colligational preference for *world* in a nested combination with SHRINK to be premodified by personal possessives. If the further environment (e.g., 15 tokens) is considered, one instance can be added as having colligational relation with the personal possessive, *her*. We can either conclude that the combination of *world* + SHRINK had a positive colligation with personal possessives or that the combination tends to be associated with a meaning of expressing PERSONAL POSSESSION.

The distinction between colligational association and semantic association in this case is not clear-cut. The borders between grammar and semantics are rarely clear because every grammatical choice is meaningful, as evidenced by our example that *world* in the combination of Personal Possessive + *world* + SHRINK was likely to be associated with the meaning of a lifestyle one may have or have had, rather than the major dictionary-cited sense of "a planet we were living on." This finding, though in need of more supporting evidence, echoes Hoey's (2014) second important observation, that when *crude* collocates with *oil* and *price*, it means *unprocessed*, but means *unsubtle* when collocating with words such as *attempt*. However, Hoey was certainly not the first to claim a correlation between meaning and pattern; as discussed in Chapter 2, Hunston *et al.* (1997) had noted co-selection between words and meaning. But perhaps it was Hoey alone who claimed that this "co-selection" was also reflected at a textual level.

World + OWE

We now examine whether the combination of *world* + OWE shared the same collocational, colligational and semantic associational features as its components and whether its more distant neighbours affected its meaning. This combines two points: whether near or far collocates associate with particular senses of *world* when combined with OWE and whether distant collocates/associations occurred with *world* and *owe* in this combination.

Coincidentally, there were also 12 instances containing the combination of *world* + OWE in this study's corpus. Excluding two cases where *world* appeared as a postmodification in a noun phrase functioning as a Subject (*What do the rich nations*

of the world owe the poor as the latter confront the issues of prevention and treatment? and Taxpayers round the rich world owe thanks to a bunch of 14 mostly-developing countries known as the Cairns group.), the remaining instances all occurred as a noun head preceding the lemma OWE and all appeared to collocate (and colligate) with the definiteness *the*. The immediate right (R1) occurrences co-occurring subsequently with this combination were *us* (two instances), *uniform calendar*, *a lot*, *you*, *each person*, and *them* (three instances). Of these collocates, *us*, *you* and *them* appeared to show that the combination of *world + OWE* had a colligational preference for accusative forms that was presumably true of *owe* on its own. However, these forms did not refer to any pronoun that might have been expected to occur earlier; on the contrary, *us*, *you* and *them* in this case referred to something vague or indefinite—e.g., those reading the texts or somebody else (e.g., *You think the world owes you. You owe the world*). Different from the combination of *world* + SHRINK, however, the five instances containing the combination of *world* + OWE + accusative forms had a tendency to express a textual semantic association with CLARIFICATION.

4) The world owes you nothing: you owe the world. Do not expect the world around you to change. It was you, in your newness, who will bring about the change.

5) We feel the world owes us. This brings a feeling of emptiness that can never be filled and leaves us with a hunger for more.

This textual semantic association with ***world* + OWE** occurred with a high LogDice score.

The previous paragraphs examined less frequent occurring examples. However, the

results shown for the two above combinations can only be convincing if evidenced by more data. Therefore, we will continue our investigation of frequently occurring combinations containing *world* and verbs following it. Two combinations, including *world + CHANGE* and *world + KNOW* were selected. The results are presented in this and the subsequent paragraphs.

world + **CHANGE**

Sixty-six percent of *world* + CHANGE (including *was changing*, *had changed*, *changed*) occurrences served as a turning point—e.g., between tradition and fashion, between old days and new days, between stubborn belief and an open mind, etc. Contexts prior to this combination (to the left) seemed to show a bad or undeveloped situation, while the contexts following the turning point (the text after the combination) were likely to show a good and developed situation—e.g., ***Our policies about employment and the economy recognize*** *that the world had changed, our principles were the same, but the world to which we're applying it* ***was very different, and, again, on that you see there would be no distinction between the so-called traditionalists and the so-called modernizers***.

world + **KNOW**

The combination of *world* + KNOW showed another kind of picture that was interesting and worth discussing. The first part of the picture was that, when we looked at the 84 instances containing *world* and *know*, we found that *the world knows* did not always mean the *world* itself knew something. Indeed, 11 out of 84 instances(13%) occurred in the combination of *in the world* and functioned as a

149

postmodification to a previous noun head, with the whole noun phrase functioning as the Subject of that clause. For example, *at least one person in the world knew him for what he really was.* Another noteworthy phenomenon, one that might be taken for granted by native speakers,   was that the semantic denotation of the word *world* was weakened in the combination of *world* + KNOW—i.e., the meaning of *world* in this cluster did not mean 'the planet we were living on' but instead served a range of functions. In the example of *Your body is perfect. Let the world know that,*[17] *the world* refers to the people living on the planet, but can also be interpreted as referring to the girl's friends, colleagues, or relatives.

Results obtaining from larger data also reveal that *world* had a particular meaning when co-occurring with particular verbs, and that the resulting word sequence (combination) was associated with a particular textual semantic association (e.g., world + OWN was associated with the textual relationship of CLARIFICATION; world + CHANGE was associated with the textual relationship of COMPARISION) when looking at their more distant neighbours.

The methodology for examining the combinations containing *world* in this chapter will also be applied to *世界 shi4jie4* to detect whether Chinese is primed between clauses or larger chunks of texts and whether the speaker-listener relationship is reflected in this kind of textual semantic association.

---

[17] *His grin flashed in the mirror. 'I can never understand people who've been brought up in your country. Why be ashamed of beauty? Your body is perfect. Let the world know that.' </p><p> She blinked in surprise and blushed with his flattery. 'Showing off seems to be the thing to do in Venice,' she retorted wryly.*

5.4.4.2 Complex issues concerning *world* + BE

Revisiting Table 5.3, which shows the most frequent R1 verbs when *world* appeared as head of the nominal group functioning as Subject, we discovered that *world* strongly collocated most frequently with BE (2,301 instances), but did not have the greatest collocational strength as calculated by LogDice score. With a relatively high LogDice and absolute high frequency of co-occurrence, the combination of *world* + BE deserves our close attention. We therefore extracted 1000 instances containing this combination to establish a database for more specific research. The instances were divided into two groups—i.e., where BE functioned as a linking verb and where BE functioned as a grammatical verb. For the first group, the typical structures were *world* + BE + adjective and *world* + BE + noun. The second group contained cases where BE served as a Passive marker (e.g., *His world was peopled by half a dozen ragged individuals*) or a Present Continuous marker (*The world was becoming a global village*). Thus, the typical structures belonging to the second group were *world* + BE + VERB+ed and *world* + BE + VERB+ing.

Statistically, 44% (440 out of 1000) of instances appeared with the structure *world* + BE + adjective, which was the strongest structure in which *world* tended to occur as a Subject collocating with BE. By contrast, only 5.5% (55 out of 1000) instances appeared with the structure *world* + BE + noun, which was the weakest structure containing *world* + BE. With regard to the second group, wherein BE served as a functional verb, we found that 33% (330 out of 1000) of instances were attributed to this structure, whereas only 17% (170 out of 1000) instances appeared with the structure *world* + BE + VERB+ing.

Adjectives following BE were scattered and hard to generalise. However, a few words formed a semantic set associated with the semantic meaning of EXPRESSING A PESSIMISTIC VIEW of the *world*, such as *vulnerable, dim, old and marginal* (13 instances).

This section has detailed the collocational and semantic associational features associated with the Subject grammatical function. Though *world* did not exhibit a clear preference or aversion for being used within this function, the number of instances provided by the data still yielded some valid findings. First, we found that *world* preferred to occur as part of a postmodification when appearing within a Subject. This was unsurprising because it seemed to be a general feature of *world* when it was compared (in Chapter 4) with other abstract nouns. The proportion of *world* appearing in the position of a postmodification was almost twice that of *use* and over twenty times that of *consequence*, and was dominantly represented by the two phrases, *in the world* and *of the world*, which were evenly distributed in our data (50:50).

Second, we found there was no clear preference or aversion for *world* occurring as individual word or in the possessive form (*world's*) as a premodification. *Worlds* modified by *world's* were found to have an association with PROBLEM, such as *pollution* and *unemployment*.

We then carried out a detailed investigation of three combinations wherein *world* occurred as noun head of the nominal group serving as Subject—two less frequently co-occurring collocations and one frequent collocation. Examination of the

combination of *world* + SHIRNK echoed Hoey's supplementary work on collocation and semantic association, which holds that a word's sense might be affected by the neighbours surrounding it, even more distant neighbours (e.g., within 15 tokens rather than five). Examination of the combination *world* + OWE showed the existence of textual semantic association. Clauses linked by this combination seemed to be associated with a meaning of CLARIFICATION.

Based on our findings from the frequently occurring combination of *world* + CHANGE, we drew a similar conclusion to those formed in our investigation of *world* + OWE—i.e., that there was a textual semantic association occurring between the clauses preceding and subsequent to this combination. Of the 61 instances containing *world* + CHANGE, 66% occurred in a context where a textual semantic association meaning of COMPARISON was implied.

5.5 Collocational profile of *world* when occurring within Complement

While *world* displayed no clear colligational relationship when occurring within a Subject or Object, it had a clear negative colligation with Complement, serving this function in only 19% (69 out of 365) of instances. However, this does not mean no interesting preference (or avoidance) was found; data derived from the instances reveal a clear preference for the use of *world's* in a premodification position within the Complement function. The ratio of *world* versus the possessive *world's* occurring as an individual modifier was 1:3, as opposed to 1:1 within the Subject and Object functions. Furthermore, words modified by the possessive *world's* constituted a colligation group of SUPERLATIVE, comprising of words such as *rarest, first,*

largest and *best*. Since the number of instances in our sample corpus was too small for more refined analysis (serving as a Complement in only 69 of the 365 instances extracted from the FLOB corpus), we turned to Sketch Engine to supplement our study. One hundred instances containing *world* within the Complement function in the BNC were selected and studied, with 73 being followed by SUPERLATIVES, 22 by the word *leading*, and five by *top*. These words together generalise a semantic association with EVALUATION OF SCALE that can be further divided into two sub-classes: NEGATIVE EVALUATION, represented by words such as *rarest, oldest, last,* etc., and POSITIVE EVALUATION, represented by words such as *leading, top, best, largest* and so forth.

Eighty-nine percent of instances appearing in a postmodification within the Complement function were realised in the form of prepositional phrase, with *in the world* occurring dominantly. However, unlike *in the world* appearing as part of a postmodification in an Object or Subject where preceding words did not permit semantic generalisation, *in the world* appearing as a postmodification within an Object had a clear preference for colligating with superlative words. Around three quarters of all instances were found to collocate with superlatives or words with a similar function, such as *first, second largest*, *the most interesting* and so on. Regardless of its grammatical position within the Complement function, there was clear evidence of *world* being used to occur with words expressing a meaning of EVALUATION, either through superlative words or similar words embracing the same sense, such as *leading* and *top*.

5.6 A summary of collocational and semantic associational behaviour of *world* within different grammatical positions serving for different grammatical functions

This section summarises the collocational and semantic associational behaviour of *world* across grammatical functions. These characteristic primings serve as a comparison base for investigating the priming features of 世界 *shi4jie4* and comparing the lexical and syntactical behaviour of the two nodes in subsequent chapters.

When appearing as noun head within an Object, verbs predicating the node word *world* were generally associated with four semantic meanings—SHAPE, GOVERN, to KNOW (the world), and ACTIONS TAKEN ON THE WORLD. The noun head had a clear preference for being associated with definiteness; the ratio between the definiteness and indefiniteness for *world* was 4:1, as opposed to around 2:1 for *consequence*, 3:1 for *aversion,* and 2:1 for *use*. While the definiteness of *world* was dominated by the article *the*, the indefiniteness of *world* at this position was dominated by *a,* with *world* being postmodified either by verb participle (e.g., *a world dominated by…*) or words linked with *of* (*a world of…*). In respect of the position of a postmodification, the prepositional phrases *in the world* and *of the world* occurred proportionally evenly (58% and 42%, respectively). In cases where *world* occurred as a premodification within an Object, it occurred individually more often than the possessive *world's*.

Verbs following *world* appearing as noun head in a nominal group that functioned as a Subject were sparse; however, they still formed a loose semantic association of

CHANGE. A stronger preference for collocating with definiteness was exhibited in instances of appearing as a noun head within an Object, with a ratio of 9:1, higher than that of other nouns except *question* (12:1).

The prepositional phrase *in the world* occupied cases occurring as part of a postmodification. The head nouns being post-modified were associated with a meaning of ORGANISATION, preceded by superlative words. The dominant structure for *world* occurring at postmodifying position was Superlative + ORGANISATION + *in the world*.

There was an avoidance for *world* appearing as a premodification when functioning as a Subject. Only 17% of instances occurred at this grammatical position, as opposed to 32% when functioning as an Object.

Having summarised *world*'s collocations and semantic associations by incorporating them with its colligational profiles, we now examine the node word's textual semantic associations. Studies on collocations, colligations and semantic associations in English have been done by numerous linguists, and it was not surprising to prove the existence of these features for the English word *world*. The data investigated answer partially the second research question—*What characteristic similarities and differences do English and Chinese display in respect of the lexical and syntactical behaviour?* The semantic associational feature of *world* was found in different grammatical patterns, which will be comparison with the semantic associations of its Chinese equivalent, *世界 shi4jie4*.

As discussed in Chapter 2, LPT has given rise to the acceptance of collocation, colligation and semantic association. What makes this theory distinctive is its attempt to explain collocation through the newly-defined notion of "priming." One of the properties of primings is that primings nest and combine; i.e., a phrase (e.g., *in winter*) has its own collocations that are separate from those of its components (*in* and *winter).* However, our finding showed the phrase *in the world* did share collocates (e.g., superlatives) with its component *world*. The only difference was that *in the world* had a stronger preference for collocating with superlatives.

5.7 Collocational behaviour of nestings containing *world*

This section examines the collocational behaviours of *world* from different angles. Its departure point is nestings; i.e., "where the product of a priming becomes itself primed in ways that did not apply to the individual words making up the combination" (Hoey 2015: 8). It then extends the span to the wider discourse to investigate the collocational and semantic features this reveals.

Sketching nestings of the node word *world* within a four-word span yielded the following top five clusters containing *world*:

> *the second world war* (3184 concordance lines)
>
> *the world cup* (1220 concordance lines)
>
> *in the third world* (675 concordance lines)
>
> *the outside world* (618 concordance lines)
>
> *in the real world* (617 concordance lines)

The need to study collocation at a greater than five-word span to its left and right has

been argued by Hoey (2014) for English and by Hoey & Shao (2014) for Chinese. Though they showed that a word's collocational behaviour was better examined in a wider context (e.g., 15 tokens), they did not examine how collocates and semantic associations differed between an English word and its Chinese equivalents.

To fill in this gap and to find evidence supporting the claim that investigation of a word's collocations and semantic association should not be restricted to a five-word span, but extended to the less immediate environment, we looked at the further left and right collocations of *the second world war*.

5.7.1 The collocational behaviours of *the second world war*

Table 5.7 below shows that, despite a relatively lower frequency of occurrence, *outbreak* ranks as the top (in terms of LogDice score) collocates within 15 words to the left of the nested combination.

Table 5.7 Left collocates of *the second world war* within 15 tokens

	Cooccurence count	logDice
outbreak	29	8.573
During	66	7.935
during	260	7.818
1939	15	7.444
Nazi	12	7.424
Cossacks	9	7.418
since	217	7.399
aftermath	11	7.349
1930s	13	7.044
Nazis	7	6.884
Depression	6	6.773
War	30	6.717
1940s	7	6.676
anniversary	10	6.611
Since	28	6.53
Germans	11	6.526
victorious	5	6.474
longest	7	6.473
repatriation	5	6.459
after	256	6.452
Hitler	7	6.393
decades	10	6.367
fascist	5	6.345
twentieth	8	6.321
interned	4	6.32
Jews	8	6.298
end	114	6.288
RAF	8	6.26
bombing	6	6.256
After	48	6.231
1920s	6	6.217
Navy	6	6.21
camps	6	6.2
advent	5	6.196
devastation	4	6.171
Welfare	5	6.148
Roosevelt	4	6.137
recession	11	6.106
1945	7	6.087
1931	5	6.078
horrors	4	6.077
Palestine	5	6.067
evacuation	4	6.061
Germany	22	6.053
characterised	5	6.034
legacy	5	6.018
wartime	5	6.011
pre-war	4	6.006
Czechoslovakia	5	6.005
Burma	4	6.003

By selecting the node combination (*the second world war*) occurring within the 29

retrieved concordance lines, we obtained wider trawls (context) for *outbreak* and *the second world war*. We then created a database for *outbreak + the second world war* by copying the wider contexts into MS Excel. For simplification, this database was named DW1.

Examining the 29 contexts with *outbreak + the second world war*, we found this word combination had a further semantic association with UP TO THE POINT IN TIME OR EVENT, in which *till/until* occurred twice, *by* four times, and *from…to…*twice. Another semantic association concerned TIME PERIOD, and was comprised of the words *before* (twice), *on* (once), *at* (four times), *prior* (once), *after* (once), and *with* (once). Looking at the context over 15 words to the left and right of the node combination, the DW1 revealed that 10 out of 29 instances (34%) had specific words for year: *1939* (four times), *1930* (twice), and *1924*, *1942*, *1907*, *1936*, and *1941* (once each).

This led us to look at the fourth collocate, *1939*. The semantic association to the left side of *1939 + the second world war* was too sparse to observe. The four cases of *1939 + outbreak + the second world war* accounted for the largest proportion (27%) of the 15 cases of *1939 + the second world war*. This combination had a strong tendency of being followed by the semantic association COUNTRIES, realised by *Europe* and *Britain*. Most know 1939 marked the beginning, or *outbreak,* in Europe of the second world war, which eventually involved more than two billion people from some 61 countries. Among these countries, America, Russia, and Britain (the Allies) comprised the main force for resistance against the Axis forces, which were mainly drawn from two European countries—Germany and Italy. This may explain

the occurrence of the terms Nazi/Nazis, Germans, Hitler, and Jews in Table 4, with relatively close LogDice scores.

In contrast to the sparse semantic associations of *1939*, the semantic associations of *Cossacks* were groupable. All nine instances of *Cossacks + the second world war* were collocated with either *repatriate* or *repatriation* within the 15-word span to the left. Four of the nine instances (80%) had *at the end of* between *Cossacks* and *the second world war.* According to Wikipedia, Cossack was a common name independently shared by several population groups and military units, the most prominent and numerous being the Ukrainian Cossacks and Russian Cossacks. Cossacks were very brave and skilful in battle, especially as cavalry. During the Second World War, they won several decisive wars at Stalingrad, which laid a solid foundation for Russia's victory in the war. Cossacks helped Russia greatly during the war, and despite being scattered throughout Eastern Europe, were treated as a minority belonging to Russia, as the historical record showed they had been living in Russia since the 17th century. They now mainly live within the territorial boundaries of Russia. This may explain why the nine instances of *Cossacks + the second world war* had a further collocate of *repatriate/repatriation*.

Then we investigated the right collocates of *world* within 15 tokens. Table 5.8 shows the general co-occurrences appearing within this span on the right side.

Table 5.8 Right collocates of *the second world war* within 15 tokens

	Cooccurrence count	Candidate count	T-score	MI	logDice
1945	12	1,816	3.457	8.909	6.871
1942	8	728	2.825	9.642	6.85
Cold	7	763	2.642	9.382	6.636
wars	8	1,413	2.822	8.686	6.47
rebuilt	6	722	2.445	9.239	6.439
Nazi	6	730	2.445	9.223	6.434
post-war	8	1,528	2.821	8.573	6.415
1940	7	1,174	2.64	8.76	6.399
Allied	6	828	2.445	9.042	6.373
1939	7	1,264	2.639	8.654	6.352
War	23	7,785	4.774	7.747	6.336
1960s	11	2,957	3.304	8.08	6.323
radar	5	588	2.232	9.272	6.264
war	47	19,261	6.817	7.471	6.21
widows	4	316	1.998	9.846	6.139
inter-war	4	359	1.998	9.662	6.106
1970s	10	3,370	3.148	7.754	6.059
1950s	7	1,944	2.636	8.033	6.039
Japanese	15	6,035	3.852	7.498	6.019
requisitioned	3	58	1.732	11.877	5.939
1930s	6	1,669	2.44	8.03	5.935
breeds	4	624	1.996	8.865	5.918
1940s	4	684	1.995	8.732	5.878
invasion	6	1,847	2.439	7.884	5.857
Stalin's	3	158	1.731	10.431	5.852
occupiers	3	179	1.731	10.251	5.834
watershed	3	208	1.73	10.035	5.81
waged	3	232	1.73	9.877	5.791
atrocities	3	233	1.73	9.871	5.79
diversified	3	241	1.73	9.822	5.783
Military	4	845	1.994	8.427	5.778
Polish	5	1,491	2.227	7.93	5.754
colonel	3	280	1.73	9.606	5.752
Attlee	3	286	1.73	9.575	5.747
nationalism	4	913	1.994	8.316	5.737
Beveridge	3	306	1.73	9.478	5.732
Allies	3	318	1.73	9.422	5.722
military	19	10,416	4.326	7.052	5.702
1949	4	991	1.993	8.197	5.692
1946	4	992	1.993	8.196	5.692
enemy	7	2,908	2.631	7.452	5.687
bombing	4	1,013	1.993	8.166	5.68
Battle	4	1,020	1.993	8.156	5.676
Nazis	3	383	1.729	9.154	5.673
Germans	6	2,352	2.436	7.535	5.657
symbol	5	1,716	2.226	7.727	5.651
scientists	7	3,033	2.63	7.391	5.647
witnessed	4	1,084	1.993	8.068	5.64
1956	4	1,095	1.992	8.053	5.634
1944	4	1,098	1.992	8.05	5.633

The semantic associations of right collocates with LogDice scores greater than 6 can be divided into three categories: TIME, represented by words expressing years, such as *1945* (12 hits), *1942* (eight hits), *1940* (seven hits), *1939* (seven hits), *1960s* (11 hits), *1970* (10 hits) and *1950* (seven hits); WAR, represented by words such as *wars* (eight hits), *post-war* (eight hits), *War* (23 hits), *war* (47 hits) and *inter-war* (four hits). Per Rychl ý, 2008, plus one point means twice the difference between collocate X and collocate Y; thus, scores lower 6 indicate a significant difference of collocational strength and should not be discussed at the same level.

The collocate candidate with the highest LogDice score was *1945*. Looking further at the cases containing *1945*, we found that, except for the sentence of *During the Second World War of 1939 to 1945, Chiswick House was occupied by the Fire Services and the house*, all samples had *1945* occurring in a separate clause or sentence. In other words, *the second world war* had a strong tendency of appearing at the end of a clause, a significant colligational feature that will be investigated in a separate chapter.

5.7.2 The collocational behaviour of *the world cup*

This section investigates the nesting of *the world cup*. We analysed this nesting as a case study to illustrate how this study's evidence can be used to affirm Hoey's argument that the grammatical category we assign to a word conventionally is "simply a convenient label we give to the combination of (some of) the word's most characteristic and genre-independent primings" (Hoey 2005: 154). This argument echoes Hoey's reversal statement on the roles of lexis and grammar, which posits that

"lexis is complexly and systematically structured and that grammar is an outcome of this lexical structure" (ibid: 1). Furthermore, we suggest, as Hoey posited in the LPT, that the word should be treated as an analytical starting point, while accepting the insights central to Sinclair's notion of the lexical item.

Below are the most frequent collocates within a -5/5 span of nestings of *the world cup* and *world cup*.

Table 5.9 Top frequent collocates of *the world cup*

Rank	Candidates	Cooccurrence Count	Score
1	finals	59	10.234
2	qualifier	23	9.476
3	qualifying	26	8.858
4	Sevens	13	8.762
5	squad	33	8.475
6	qualify	16	9.949
7	Murrayfield	6	7.602
8	Wallabies	5	7.426
9	Final	8	7.415
10	quarter-final	5	7.273

Table 5.10 Top collocates of *world cup*

Rank	Candidates	Cooccurrence Count	Score
1	finals	109	10.248
2	qualifying	109	10.138
3	qualifier	76	10.026
4	1995	64	9.401
5	squad	87	9.349
6	Rugby	50	8.971
7	England's	54	8.938
8	Sevens	24	8.409
9	Scotland's	33	8.371
10	tie	48	8.353

The semantic associations of *the world cup* were GAME TERMINOLOGY (e.g., *finals, quarter-final*), TEAM MEMBERS (e.g., *Wallabies*, [Rugby] *sevens*), while those of *world cup* were TIME (e.g., *1995, 2007*), TEAM (*squad, tie, match*), and TEAM NAME (e.g., England, Scotland), indicating the collocational behaviours of the two differed in terms of semantic associations. Though sharing some collocates, the strength of attracting these candidates varied. For example, the strength of *qualifying* occurring with *the world cup* was 8.858, whereas the strength of it occurring with *world cup* was 10.138. As discussed previously, a difference of only two scores represents a large difference in collocational strength; in other words, the difference in LogDice scores for *qualifying* between *the world cup* and *world cup* reveals the latter nesting had a stronger attraction for the word *qualifying*.

The difference between the two combinations of *the world cup* and *world cup* was only the word *the*, which was seen as an "obligatory lexis" to the noun phrase *the world* by Stubbs (2008:165), who considered *the world* and the combination of *world cup* to be a lexical unit, or a "partly fixed lexical core plus other variable items" (ibid: 163), and argued that knowing the meaning of the components did not tell you the meaning of the whole combination. Stubbs (2008) held that the study of a term's phraseology should be extended to include the study and discussion of its combinations of collocation, colligation, and other evaluative communicative functions. His point of view was discussed in the literature review in reference to Sinclair's "unit of meaning" model. We accept his and Sinclair's insight that collocation is not "mere juxtaposition" (Firth cited in Palmer 1968: 181), but rather "recurrent phrasal constructions [or nesting in LPT]" (Stubbs 2007:163); however, we suggest we had best not "rush too quickly to close off the upper boundary of the

lexical item" (Hoey 2005: 158).

Following Stubbs's line that *world cup* is a lexical item with a new meaning different from the intersection of its parts, this lexical item should show a similar collocational behaviour as *the world cup*, as the only difference between the two is the obligatory lexis, "the." However, data extracted from the BNC paint a different picture. Sinclair's (2004a) terms "a single choice" and "a single lexical time" are not identical as a word and word sequence in the LPT, though the concept of primings also applies to longer sequences (nestings) built out of individual words. Hoey tackled this issue head-on by underlining that priming have a more "textual dimension" (Hoey 2005: 158), "with networks of primings informing considerably longer stretches of text than Sinclair had in mind" (Stewart 2010: 156). Hoey claimed that "priming contextualizes theoretically and psychologically Sinclair's insights about the lexicon" (ibid) in operating nesting in a systematic way for "the move from the word to the lexical item, and from the lexical item to the wider text" (ibid).

Hoey's claim can account for results obtained from *the world cup* and *world cup*. The word *world* was first primed to collocate with *the*, after which *the world* had a strong attraction to *cup*. This combination in turn had other sorts of collocates, as displayed in Table 5.9. Similarly, we can account for the collocational behaviour of *world cup* as follows: the word *world* is primed to collocate with *cup* and this combination in turn has other sorts of collocates.

The different primings of *the world cup* and *world cup* can also be accounted for by

their colligational differences. Of the 26 instances of *the world cup qualifying*, 81% of cases (21 out of 26) occurred as premodifiers in a noun phrase functioning as an Object (e.g., *Craig Brown will see out the World Cup qualifying series*). In contrast, of the 109 instances of *world cup* qualifying, 53% (58 out of 109) occurred as premodifiers in a noun phrase functioning as an Adjunct (e.g., *Robson's team had yet to concede a goal in their five World Cup qualifying matches*). Nestings of *the world cup* and *world cup* tended to belong to the grammatical category of ADJECTIVE, a feature neither of its components seemed to have. Even the part of speech of *world* in *world cup* and *the world cup* was worthy of exploring in detail. Hoey suggested calling the grammatical category to which a word belongs its "grammatical priming" (Hoey 2005: 155), and further claimed that "if we accept that grammatical categories were labels for combinations of primings, we had also to accept that the primings of some words or word sequence will not permit the application of the conventional grammatical labels (or any labels)" (ibid: 156). As discussed in the literature review, Sinclair undertook an investigation of the word *of* in 1991, which he reiterated in 2004. The significance of his study on the conventionally viewed preposition was that its collocations and colligations were substantially different from those of other words labelled as "preposition."

Sentences containing words primed differently from their conventional grammatical role were extracted by Hoey as follows, with the relevant words emboldened:

6) The expedition returns to England, having rescued the men left to **winter** on Elephant Island and picked up the party from McMurdo Sound. [verb]

7) Many happy **returns**… [noun]

8) I found it growing here and in more northerly outlets behind a sea wall, heavily **picked** but with little sign of exploitation [adjective]

9) There's less to **party** about [verb]

10) The man **pockets** his hands. [verb, from O. Henry's *After Twenty Years*]

(Hoey 2005: 155-156; last instance added by author)

The research results of the two aforementioned combinations attest to Hoey's argument on reversing the traditional view of grammar as systematic and lexis as loosely organised. The conventional category "noun" (or "verb" or "adjective" or others) "gifted" to the nested combinations can itself be colligationally primed when occurring for other words.

5.8 Summary of collocation and semantic association of *world*

This chapter has demonstrated evidence of characteristic primings with respect to *world*'s collocation, colligation, and semantic association. In addition to applying traditional LPT, we extended our study to the co-textual level and investigated the interplay between the three concepts. Our findings can be summarised as follows:

1. The collocations and semantic associations of a nesting containing the node *world* may differ from those of its components; however, these differences are neither extreme nor absolute. The node *world* and nesting containing it can share the same collocates and semantic associations, while differing in the strength thereof. For example, both *world* and the combination of *in the world* collocated with the

semantic association of SUPERLATIVE, but *in the world* had a stronger tendency of collocating with this semantic association than *world* and the semantic sets substituting to this semantic association were far more centralised than those of *world*. *World* was primed to occurred with *the*, and *the world* was in turn primed to occurred with *in* to form a combination of *in the world*. This combination's collocations and semantic associations, however, differed from its components', as reflected in their different collocates and semantic sets; when collocates and semantic sets were shared, they differed in collocational strength.

2. Semantic association sometimes may have only one corresponding collocation, such as *world* + *owe*; among the verbs subsequent to *world*, no other similar semantic set formed this semantic association. That is to say, *owe* formed a semantic association with only itself. Average corpora are unlikely to detect semantic sets of this semantic association further and may fail to provide sufficient statistical evidence. The inability of corpora shows that primings differ from person to person and may move from collocations to semantic associations. Thus, it is not necessary for a corpus to reflect collocations that lead to a particular semantic association, since they might be primed for only a few speakers in only a few contexts.

3. Semantic associations are grammatically dependent. The word *world* had a semantic association of CHANGE, but CHANGE was not always provided in the same structural form:

    11) The technology has *changed the world* so much in last century.

    12) The *world is changing*…

The sequence *in the world* was associated with SUPERLATIVE, again with some variation in the wording of the SUPERLATIVE:

13) British supermarkets are the *best* of their kind *in the world*.

14) He was once described as the *best in the world*.

In short, the semantic associations of both a word and a word sequence it makes up appear to be grammatically tied. The possibility of a word being primed under certain grammatical conditions without being an implication of semantic association will also be explored for Chinese in subsequent chapters.

4. It is suggested that the semantic features of a word need to consider greater spans, as more distant neighbours may show semantic associations similar to those of the more immediate neighbours. For example, we found that the combination of *the second world war* tended to have a semantic association of TIME within the broader environment (e.g., 15 tokens) and *the world cup* tended to associate with TEAM MEMBERS in an extended context. There are reasons to believe that a word's neighbours may strongly affect the senses of the word in question and that the word's relationship with its neighbours can be extended to a greater span.

5. Preliminary evidence suggests that semantic association and pragmatic association are two sub-classes of semantic prosody. Pragmatic association is not always consistent with semantic association. For example, the combination of *world + shrink* had a semantic association with TECHONOLOGY and the potential of expressing a negative point of view, since the word *shrink* inherits a negative sense.

However, writers in the BNC applied this combination to simply state the timeless truth that the world has been changed by the introduction of advanced technology, such as telephones and the Internet. The textual association of stating a timeless truth was also represented by a set of words, such as *enter* and *go into*. In addition, our evidence suggests that nested combinations of words are more likely to be primed for textual semantic association than the individual word in question, perhaps because the word sequence can be seen as a priming product of various collocational, semantic associational primings.

Chapters 4 and 5 have investigated the three essential and fundamental concepts of LPT for the English node word *world*. In subsequent chapters, we will shift our attention to the investigation of Chinese to explore whether the lexical and grammatical behaviours found for *world* exist for its Chinese equivalent, *世界 shi4jie4*.

CHAPTER 6

Collocations and Colligations of 世界 *shi4jie4*

6.1 Introduction

In Chapter 4 and Chapter 5 we presented evidence related to the collocational, colligational and semantic associational behaviour of *world* and showed that *world* or nestings containing *world* preferred or avoided occurring within certain grammatical patterns, sentence positions, or even specific types of semantic relationships. The data confirm that priming operates on both individual words and word sequences and that a word or word sequence can also be primed for textual relationships. We also made the tentative claim that the definition of *nesting* in LPT may need modification. As a product of a word sequence, nesting becomes itself primed but also shares some collocates or even semantic associations with its components. The difference lies in the strength of co-occurrence with other words in certain contexts. In our study, this kind of strength could be calculated using LogDice value. Following Hoey, we also proposed, based on the findings in the previous two chapters, that the length of studying a word range's collocation could be expanded, rather than restricted to a four- or five-word span. The span set by Sinclair (1991) was an appropriate length for determining the sequence of a nesting, but not long enough for determining a word's collocational behaviour, which suggests that priming works for a longer sentence or even a particular context or domain (genre).

In this and the subsequent chapters, we seek to detect whether LPT also applies to the

Chinese node 世界 *shi4jie4*. Collocation and colligation were considered in this chapter, while complex issues combining collocation, colligation and semantic association will be discussed in next chapter. By examining data retrieved from the corpora, we aim to answer the first and main question of this present study—i.e., *1.Can collocation, colligation and semantic association, as defined in LPT, account for the range of lexical and syntactical behaviour in Chinese?* With the collocational, colligational and semantic associational behaviour shown by 世界 *shi4jie4* and word sequences containing 世界 *shi4jie4*, we will also make a comparison between *world* and 世界 *shi4jie4* to detect their similarities and differences.

This chapter adopts the following conventions of representation:

1. The representation of a Chinese lexical item was written with its character (字 *zi4)* in the initial position, followed by its Pinyin, the tones of which were written in Arabian numbers (from 1-4). The word's English equivalent follows in brackets. For example, the Chinese equivalent of *world* was represented as 世界 *shi4jie4* (world).

2. The representation of examples extracted from Chinese corpora was displayed with Chinese on the top, followed by the Pinyin to each character. A word-to-word English translation was then provided, followed by a readable translation. For example,:

中国	是	当今世界上	最大的
zhong1guo2	shi4	dang1jin1shi4jie4shang4	zui4da4de
China	was	today's world	the largest

建筑　　　　　王国。

jian4zhu4　　　wang2guo2.

construction　　　kingdom.

China is the largest construction kingdom in today's world.

6.2 The general collocational behaviour of 世界 *shi4jie4*

There were 1,079,563 instances of 世界 *shi4jie4* occurring in the zhTenTen11 corpus, which far exceeds the occurrences of *world* in the BNC (58,496). This was simply attributable to the larger size of the Chinese corpus used in the study, however; both occurred with similar frequency per million running words—specifically, 世界 *shi4jie4* occurred 512.45 times per million running words, while *world* occurred 520.94 times per million running words. This similarity shows that speakers of both English and Chinese use the word *world* or 世界 *shi4jie4* with high frequency. Stubbs (2007: 164) explained that "the word *world* was frequent because in which it occurred frequent phrases." This word frequency has been noted in previous chapters, where we discovered *world* was likely to occur in phrases such as *in the world*, *all over the world*, *the rest of the world*, and so forth. We need now to account for the equally high occurrence of 世界 *shi4jie4* in Chinese corpus. We assume that 世界 *shi4jie4*, like its English equivalent, also tends to occur more often in phrases, such as 世界中 *shi4jie4zhong1*, 世界上 *shi4jie4shang4* or 全世界 *quan2shi4jie4* (all these combinations can be translated as *in the world*, although the last also has such alternative translations as *all over the world* or *the whole world*).

As we did for *world*, we first retrieved concordance lines containing the node within a five-word span to both the left and right sides of 世界 *shi4jie4*. The collocates were sorted in descending order based on their LogDice score (see Table 6.1 below). The L1 position was dominated by two major parts of speech—adjectives and verbs—with the former including 全 *quan2* (all), 当今 *dang1jin1* (today), and 整个 *zheng3ge4* (whole), etc., and the latter including 走向 *zou3xiang4* (walk to), 成为 *cheng2wei2* (become), 居 *ju1* (rank), 来自 *lai2zi4* (come from), and 占 *zhan4* (take or account for), etc..

Table 6.1 L1 collocates of 世界 *shi4jie4*

Rank	Collocate		LogDice
1	全	all	8.799
2	当今	today	8.579
3	走向	walk to	8.127
4	成为	become	8.105
5	这个	this	7.953
6	居	rank	7.589
7	来自	come from	7.318
8	乃至	and even	7.231
9	整个	whole	7.191
10	届		7.120
11	占	take	7.021
12	内心	mind	6.943

At the L2 position (see Table 6.2) there were some shared collocates with L1, falling into two main grammatical categories—VERBs and ORDINAL WORDS—with the exception of the definiteness marker 这个 zhe4ge4 (this).

Table 6.2 L2 collocates of 世界 shi4jie4

Rank	Collocate		LogDice
1	成为	become	8.456
2	走向	walk to	8.208
3	这个	this	8.055
4	乃至	and even	7.656
5	居	rank	7.605
6	来自	come from	7.516
7	向	toward	7.290
8	第二	second	7.002
9	面向	turn towards	6.969
10	第三	third	6.926
11	在	in/at/on/around	6.861
12	是	BE	6.810

The verb group included 成为 cheng2wei2 (become), 走向 zou3xiang4 (walk to), 乃至 nai3zhi4 (and even), 居 ju1 (rank), 来自 lai2zi4 (come from); and the ordinal words included 第二 di4er4 (the second) and 第三 di4san1 (the third). In addition to these two major categories, there were also two typical collocates which could be equated to several lexical items in English, even if written in only one character in Chinese. The first typical collocate was the prepositional word 在 zai4, which is equivalent to many prepositions in English, such as *in, at, on and around*, depending on the context it appears in. Another typical collocate was 是 shi4, an equivalent to the BE word group in English.

We then investigated collocates at the L3 position to 世界 *shi4jie4* and found they

seemed to be associated with a semantic meaning of CHANGE (e.g., 改造 *gai3zao4*

(transform) and 改变 *gai3bian4* (change), and REGION (e.g., 中国 *zhong1guo2*

(China)*, 我国 *wo3guo2* (our country), and 西方 *xi1fang1* (Western).

Table 6.3 L3 collocates of 世界 *shi4jie4*

Rank	Collocate		LogDice
1	这个	this	8.195
2	中国	China	7.685
3	我国	our country	6.824
4	已	has	6.792
5	让	let	6.703
6	作为	as	6.566
7	拥有	own	6.526
8	认识	to know	6.519
9	对	for	6.501
10	改造	transform	6.469
11	改变	change	6.460
12	西方	Western	6.366

In a comparison with the results from Chapter 4, the data displayed in above tables

show there were some collocates shared by English and Chinese to the left side of

both *world* and 世界*shi4jie4*. At the L1 position, *world* and 世界*shi4jie4* shared the

collocates *whole*/全 *quan2* (or 整个 *zheng3ge4*) (whole) and *today's*/当今 *dang1jin1*

(today's). The slight difference was that *today's* had a lower collocational strength of

co-occurring with *world* and thus was not distributed in Table 4.1; however, it also

belonged to the semantic association of POSITIONING IN TIME of *world*, which

was represented by more frequent collocates such as *modern* (appearing in Table 4.1), which could also be an optional equivalent to Chinese 当今 *dang1jin1* (today's), depending on the context in which it occurred.

Table 6.1 does not list every frequent collocate at the L1 position to 世界 *shi4jie4* due to the allowance of pages. Further down in the list produced by Sketch Engine, collocates co-occurring with 世界 *shi4jie4*, though appearing with lower LogDice scores, could also be grouped into a semantic set forming a semantic association of POSITIONING IN TIME, such as 目前 *mu4qian2* (at present) (6.919) and 当时 *dang1shi4* (at that time) (4.996). These semantic set members did not occur at the L1 position for *world*. This comparison shows that *world* and 世界 *shi4jie4* shared the same semantic association at the same position, although they differed in terms of the members constituting the semantic association.

Now we consider the second position to the left side of *world* and 世界 *shi4jie4*. Table 4.2 reveals that the L2 position of *world* was dominantly occupied by prepositions, represented by *around*, *throughout*, *outside*, *over*, *in*, etc. This phenomenon also occurred at the L2 position of 世界 *shi4jie4*. As mentioned earlier, Chinese is not a language with many variants of prepositions. The preposition 在 *zai4* at L2 to 世界 *shi4jie4* could be treated as equivalent to different prepositions in English. From this finding, we can safely conclude that 世界 *shi4jie4,* in the same way as *world,* had a strong colligational preference for occurring with prepositions at the L2 position.

In addition to the colligational behaviour of co-occurring with prepositions, we discovered that 世界 *shi4jie4* was likely to occur with ordinal items, such as 第二 *di4er4* (second) and 第三 *di4san1* (third), at the L2 position. These lexical items occurred at the L2 position to 世界 *shi4jie4* rather than at the L1 position (as they did with *world)* simply because the classifier 次 *ci4* occurred in between. In Chinese, 第二次 *di4er4ci4* and 第三次 *di4san1ci4* are fixed co-occurrences that correspond to the word combinations *the second* and *the third,* respectively, in English.

It has been suggested by Wei and Li (2014) that cross-linguistic equivalence resides in corresponding patterns of units, not corresponding single words. Our examination echoes this view and further show that the equivalence between English and Chinese is not restricted to the number of tokens. As discussed in Chapter 3, in a cross-linguistic study between English and Chinese, it is ultimately important to define what 'word' means before carrying out the examination. Tying the notion of 'word' to a written symbol (here referring to Chinese characters) is somewhat arbitrary. We attempted to resolve this problem by applying the definition of nesting proposed in LPT. For English, Hoey defined nesting as a product of primings manifested by individual words (the components of a nesting). For Chinese, we can follow this definition and consider a Chinese 'word' to be a nesting made of monosyllabic characters or morphemes. Thus, the word *second* in English can be equated to the nesting of 第二次 *di4er4ci4* in Chinese. By doing so, we no longer need to work out an equivalent definition for *word* as linguists have done for English, as we were comparing English and Chinese from various perspectives. Pedagogically, the notion of nesting reduces some memory work and tends to be in accord with the perceptions of most people — particularly English-speaking students

attempting to learn Chinese. In the next chapter, the assumption of nesting's validity for Chinese will be tested and explored by integrating colligational and semantic associational profiles of 世界 *shi4jie4*.

We now turn to the right side collocates of 世界 *shi4jie4*. The most frequent R1 collocates were generalised into semantic associations (see Table 6.4), for example, FAMOUS represented by lexical items such as 公认 *gong1ren3* (recognised by the public), 著名 *zhu4ming2* (famous), 闻名 *wen2ming2* (renowned), 知名 *zhi1ming2* (well known) and FIRST CLASS, represented by 一流 *yi1liu2* (first), 领先 *ling3xian1* (leading), 顶级 *ding3ji2* (the best, the top) and *500* (indicating the top 500 enterprises in the world). Combinations made up of these items and 世界 *shi4jie4* served as part of premodification in a nominal group. This colligational behaviour will be discussed in detail later in this chapter.

Table 6.4 R1 collocates of 世界 *shi4jie4*

	Rank	Collocate		LogDice
1		一流	first	8.195
2		公认	recognised	7.685
3		大战	war	6.824
4		上	up	6.792
5		著名	famous	6.703
6		闻名	renowned	6.566
7		和平	peace	6.526
8		顶级	top	6.519
9		末日	end	6.501
10		知名	well known	6.469
11		领先	leading	6.460
12		500		6.366

Unlike collocates occurring at the R1 position for *world*, the top R2 collocates of 世界 *shi4jie4* demonstrated more variants and shared candidates with collocates occurring at the R1 position, which were dominantly made up of adjectives, such as 一流 *yi1liu2* (first), 公认 *gong1ren4* (recognised), 著名 *zhu4ming2* (famous), 闻名 *wen2ming2* (renowned) and so forth (shown in Table 6.5). These candidates were members of a semantic set forming semantic associations of FAMOUS and FIRST CLASS. In addition to these shared collocational members, words co-occurring at the R2 position were likely to be colligated (and also collocated) with SUPERLATIVE, represented by 最 *zui4* (most) *and* 最早 *zui4zao3* (earliest).

Table 6.5 R2 collocates of 世界 *shi4jie4*

Rank	Collocate		LogDice
1	一流	first	8.015
2	公认	recognised	7.891
3	最	most	7.602
4	著名	famous	7.344
5	唯一	only	7.3
6	大战	war	7.255
7	上	up	7.164
8	最早	the earliest	7.038
9	遗产	heritage	6.58
10	和平	peace	6.506
11	顶级	top	6.449
12	闻名	renowned	6.401

Table 6.6 below reveals that while the R3 position of *world* exhibited fewer variants, collocates occurring at the R3 position of 世界*shi4jie4* were more varied, dominantly

occupied by words such as 经济体 *jing1ji4ti3* (economic community), 强国 *qiang2guo2* (powerful country), and 工厂 *gong1chang3 (factory)*. These collocates also showed a strong tendency to be associated with a semantic meaning of ECONOMY (e.g., 经济体 *jing1ji4ti3* (economic community) and 贸易 *mao4yi4* (trade).

Table 6.6 R3 collocates of 世界 *shi4jie4*

Rank	Collocate		LogDice
1	上	Up	7.179
2	最高	highest	7.131
3	潮流	trend	6.583
4	古老	ancient	6.392
5	经济体	economic community	5.604
6	强国	powerful country	5.542
7	工厂	factory	5.259
8	规模	scale	5.214
9	锦标赛	championship	5.183
10	贸易	trade	5.081
11	文明	civilisation	5.029
12	难题	problem	5.017

Tables 6.2-6.6 display the most frequently co-occurring collocates at the right side of 世界 *shi4jie4*. The tables show that 世界 *shi4jie4* had a wider range of collocates occurring at positions of R1 to R3, in comparison of collocates occurring at the same positions of *world*. Nevertheless, *world* and 世界 *shi4jie4* shared a few top frequent collocates. At the R1 position, they shared the collocate of *war/*大战 *da4zhan4* and at R2 they shared the SUPERLATIVES *most* (最 *zui4*) and *biggest* (最大 *zui4da4*), where the latter was also a shared colligational behaviour.

Having detected and sorted out the collocates at each position to the noded item 世界 *shi4jie4*, we now sketch a general picture of the collocates occurring within a five-word span at both sides of 世界 *shi4jie4*. Left and right collocates occurring at each position will be sorted out as a whole and results displayed in Tables 6.7 and 6.8, respectively, in descending order of LogDice score.

Table 6.7 Frequent left collocates of 世界 *shi4jie4*

Collocates		Frequency	LogDice
全	all	64,312	8.935
成为	become	30,393	8.761
当今	today	14,451	8.69
中国	China	51,428	8.4
这个	this	25,689	8.378
走向	walk to	12,056	8.357
世界	world	20,830	8.304
目前	current	14,917	7.826
乃至	until	7,516	7.724
来自	from	8,826	7.663
居	stand	7,327	7.662
整个	whole	8,127	7.531
美国	America	10,947	7.526
届		10,251	4.678
向	toward	14,906	4.158

Table 6.8 Frequent right collocates of 世界 *shi4jie4*

Collocates		Frequency	LogDice
一流	first level	31,190	9.698
最	Most	59,219	9.154
上	Up	117,185	9.069
国	Country	33,527	8.73
遗产	Heritage	12,967	8.517

经济	economy	40,608	8.462
大战	war	11,841	8.454
各	every	59,637	8.414
世界	world	20,830	8.304
第一	No.1	23,045	8.289
著名	famous	11,938	8.213
范围	range	14,221	8.187
先进	advanced	14,030	8.155
大	large	46,146	8.149
水平	level	18,526	8.127

In Table 6.7, the literal English translation for the penultimate collocate is blank. This is because the word 届 *jie4* is a classifier in Chinese and has no word-to-word English equivalent. In English, this classifier is mostly expressed by an ordinal number. For example, 第一届奥林匹克运动会 *di4yi1jie4 ao4lin2pi3ke4 yun4dong4hui4* (the first Olympic Games), where 届 *jie4* is omitted in the process of corresponding to its English counterpart.

The general collocational picture drawn in this section shows that at each position, the node words 世界 *shi4jie4* and *world* shared some, even if few, collocates and semantic associations. As with the semantic associations of *world*, the semantic associations of 世界 *shi4jie4* were also made up of both high- and lower-frequency collocates. Priming, as proposed in LPT, can account for this phenomenon, in that priming moves outwards from collocations to semantic association. It is probable that collocations were primed first and that semantic associations then formed with the collocations to produce more abstract priming, whether as self-reflection or through encounters, to reinforce the more general priming. The order might be reversed; i.e., the general semantic priming may be formed on a semantic set, thus

permitting creative word choices to be made.

So far, we have investigated the general collocational behaviour of 世界 *shi4jie4* and compared the most frequent collocates of 世界 *shi4jie4* and *world* at different positions. However, a holistic picture of 世界 *shi4jie4* cannot be drawn without considering its colligational profile by adopting a co-colligational approach. The next section explores the colligational behaviour of 世界 *shi4jie4* to test whether the concept of colligation as defined in LPT is also applicable to 世界 *shi4jie4* and, if so, how colligation may operate for Mandarin Chinese.

6.3 Colligational profile of 世界 *shi4jie4*

Before providing a comprehensive picture of the colligational behaviour of 世界 *shi4jie4,* we must first demonstrate whether colligation, as defined in LPT, is also valid for Chinese. From a psychological perspective, the notion of colligation in LPT features in the preference or avoidance a word or word sequence embraces. Repulsion or aversion does not mean a word or word combination tends to occur at a particular place within a group, serve a particular function in a clause, or be positioned at a particular place in a sentence; rather, it reflects an individual's intention when applying the word or word sequence. The kinds of colligational features accumulated from each individual's use are shown in a corpus made up of the language used by speakers of a particular community. Although a corpus represents no one person's experience, it may be analogous was some respects to in individual's experience due to its ability of concordancing a collection of

accumulated instances of a piece of language.

In this section, we will apply a corpus-driven approach to detect whether 世界 *shi4jie4* showed a preference or avoidance for occurring within a lower group and in a higher rank.

6.3.1 General colligational profile of 世界 *shi4jie4*

In pursuit of the potential colligations of 世界 *shi4jie4,* we examined 535 instances containing 世界 *shi4jie4* drawn from the LCMC corpus. Following the sequence, this database was numbered as DW 6. As with English, a Chinese noun can also be part of some group or other word sequence and in turn that group or word sequence performs some function in a clause. The procedure is similar to that used in investigating *world* in Chapter 4. First, we examined colligations of 世界 *shi4jie4* in the clause and then looked at its characteristic colligational behaviour within a nominal group. Second, we investigated the colligations of 世界 *shi4jie4* when occurring in a phrase, such as 世界上 *shi4jie4shang4* (in the world) and 在世界范围内 *zai4shi4jie4fan4wei2nei4* (in the range of the world), etc.

6.3.1.1 The distribution of 世界 *shi4jie4* occurring in genres

The proportion of 世界 *shi4jie4* distributed in different genres of the LCMC was

calculated as follows:

- Step 1: We selected the first five texts from each genre and counted the number of clauses of each selected text.

- Step 2: We counted the average number of clauses across these five texts. Then we multiplied these averages by the total number of texts of each genre to get an approximate number of clauses in every genre as a whole.

- Step 3: We retrieved instances containing 世界 *shi4jie4* from each genre. These instances were then divided by the total number of clauses of the genre they appear in to calculate the proportion of 世界 *shi4jie4* cases in all clauses.

Proportions lower than 0.1 were not included in the chart, such as Genre K (General Fiction). The proportional distribution of instances containing 世界 *shi4jie4* across genres in the LCMC is depicted below:

Figure 6.1 Distribution of 世界 *shi4jie4* occurring in genres of the LCMC
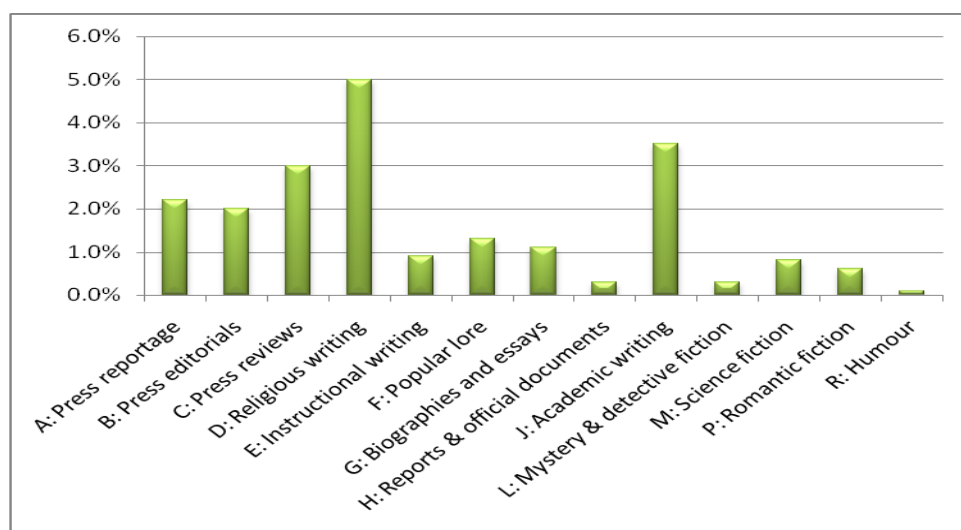
As can be seen from the chart, *世界 shi4jie4* occurred proportionally most often in Genre D: Religious writing; an example extracted from texts of this genre is illustrated in Example 1. Academic Writing had the second largest proportion of instances of *世界 shi4jie4*, as illustrated in Example 2. The third most common genre in which *世界 shi4jie4* tended to appear was Genre C: Press review, an example of which is shown in Example 3.

1)	佛教	在世界各地	都有人
	fo2jiao4	zai4shi4jie4ge4di4	dou1you3ren2
	Buddhism	in the world everywhere	there were people
	在研究、	在信仰、	在实践。
	zai4yan2jiu1	zai4xin4yang3	zai4shi2jian4
	being studied	being believed	being practiced

Buddhism is being studied, believed and practiced in every corner of the world by many people.

2)	90 年代	世界政治	是
	90nian2dai4	shi4jie4zheng4zhi4	shi4
	In 1990s	world politics	was
	复杂	而	多变的。
	fu4za2	er3	duo1bian4de
	complex	and	uncertain

In 1990, the political [situation] in the world was complex and uncertain.

3)	来自		六个不同国度的

lai2zi4	liu4ge4bu4tong2guo2du4de
From	six different countries
六名考察队员	成为
liu4ming2kao3cha2dui4yuan2	cheng2wei2
six expedition members	become
世界各国	人民
shi4jie4ge4guo2	ren2min2
every country in the world	people
心中的	英雄。
xin1zhong1de	ying1xiong2
in the heart	hero

The six expedition members were from six different countries. They have become heroes in the hearts of people from every country of the world.

Figure 6.1 also indicates that, in addition to the three genres with the majority instances, 世界 *shi4jie4* also favoured the first two genres—Press reportage (Genre A) and Press editorials (Genre B)—with proportions of 2.2% and 2%, respectively. These two genres, along with Press Reviews, were categorised as a register of NEWS in the LCMC. Thus it is safe to conclude that 世界 *shi4jie4* had a strong preference for occurring in the NEWS domain.

On the other hand, instances containing 世界 *shi4jie4* were rarely found in Genre R: Humour, with only two instances appearing in approximately 450 clauses. The characteristics of 世界 *shi4jie4* occurring in genres of the LCMC show that, as with English, a Chinese word may also prefer or avoid appearing in a particular kind of

writing.

6.3.1.2 Tense of 世界 *shi4jie4*

We discussed in Chapter 3 that Chinese, unlike English, does not have a systematic tense system reflected in its verb forms, such as *look → looked* or *go → went.* Chinese learners of English as a foreign language, especially at the beginning level, are taught the importance of tense in English; they are told, for example, that *looked* is the Past tense of *look*, so that when you "look" at something yesterday you must change its word form to *looked*. It is significant that some writers on Chinese have denied altogether that Chinese contains any grammatical form of expressions of time; for instance, Maspero (1937:6; quoted and translated by Halliday 2007: 117) wrote: "what grammatical forms there were expressing subjective attitude rather than any definite system of time-relation." However, a somewhat different view of Chinese was held by earlier philologists, such as Saussure, who contrasted Chinese "apparent formlessness with the transparently grammatical structure of the Indo-European languages" (quoted in Halliday 2007:118). Saussure called Chinese "ultra-lexical," referring especially to classical Chinese, where grammatical relations are expressed through lexical items, word order and so on.

This insistence on the absence of strict formal grammatical categories in Chinese has been maintained for a long time, leading some linguists, including both Chinese linguists (e.g., Chao 1945) and Indo-European language linguists (e.g., Frei 1941), to point out the danger of arbitrarily applying the grammatical categories of an

Indo-European language (such as English) to Chinese. Nevertheless, many writers of early textbooks on Chinese tried to mould Chinese verbs into a tense-system based on or taken directly from that of Latin. These attempts, despite not having a clear rational basis, still survive in modern works, such as Mullie's tense-scheme of the Chinese verb (1970).

According to Halliday (2007: 203), "Tense was temporal location of an abstract, non-lexical (semantically grammatical) kind, relating only to the time of speaking (primary tense) or (secondary tense) to the time spoken of"; thus, it has no need to be expressed in a tense-form. Halliday undertook an examination of the forms of expressions of temporal categories in Chinese by comparing them with particular modes of expression in particular languages, such as aspect in Slavonic, to ascertain how far and in what way the definite temporal structure is expressed in modern Standard Chinese; specifically, the spoken language of Beijing. He concluded by arguing with Chao—who refused to use "tense" as a formal category—that the tense system cannot be associated with one kind of distinction, and that Chinese contributes to an "extra distinction" (Halliday 2007: 203), with a neutral (semantically unmarked) category expressed by neutral (formally unmarked) form. Halliday's examination of the suffixes (or Chinese morphemes used by Li and Thompson 1989) 了 *le* and 来着 *lai2zhe1* showed that when the former is added to the predicate as a whole (e.g.我吃饭了 *wo3chi1fan4le I had dinner*) and the latter is added to the verb itself (e.g.我跑步来着 *wo3pao3bu4lai2zhe1 I ran*), they tend to have a Past meaning.

In addition to these suffixes, Chinese—like English or other Indo-European

languages—also has adverbs indicating a clearly temporal category. For example, 他
过去是一名士兵 *ta1guo1qu2shi4yi4ming2shi4bing1 He was a soldier in the past.*
过去 *guo4qu4* (in the past) in this sentence is a characteristic marker equivalent to a
Past tense in English.

However, the Future tense in Chinese, like the Future tense in English, may not show
a "real" future activity taking place in certain days/months/years later. For example,

4) 他　　　预料　　　世界　　　　将会

　　ta1　　yu4liao4　shi4jie4　　jiang1hui4

　　He　　predicts　world　　　will

　　发生　　　　　　另一次　　　大灾难。

　　fa1sheng1　　　ling1yi1ci4　da4zai1nan4

　　happen　　　　anther(another?) disaster

He predicts that there will be a big disaster occurring in the world.

In the above example, the lexical items 将 *jiang1* and 会 *hui4* are typical markers for
a Future tense in Chinese; however, the meaning reflected from sentences containing
either of these words does not always show a truth that will definitely happen in a
certain time in the future, but a prediction of or assumption about a possible disaster
that may happen. English also has this kind of ambiguous usage; for example, *He
will get married* may mean that the man will get married on a particular day in the
future or it may just show the speaker's confidence that the man will find a wife in
his lifetime.

As in English instances containing *world*, the analysis of Future tense in Chinese was

also unproblematic in this project, because the tiny number of occurrences of Future 'tense' (five out of 535 instances) did not entail any characteristic that could be used to support any part of the ambiguity aroused by the use of Future tense.

The above discussion of tenses and aspects shows that Chinese does not manifest as high a degree of grammatical morphemes as does English. However, the languages are comparable in terms of tenses and aspects, because Chinese has markers for signalling the time and duration of a reported event either relative to the time of speaking (tense) or other events (aspect). It is not problematic to analyse the Present and Past tense since the makers are rather obvious and it is not that demanding to infer the tense based on the context.

As was previously done in Section 5.1, we calculated a normalised distribution of the LCMC in advance to determine whether 世界 *shi4jie4* had a preference/avoidance for a specific tense use. If the preference of its occurrence was in connection with a particular tense that corresponded to that of the normalised tense distribution of the LCMC, the preference was not colligationally interesting; however, if the preference contradicted or outweighed the normalised distribution to a considerable extent, the feature was colligationally interesting and worthy of attention.

The procedure for obtaining the normalised distribution of the LCMC was the same as that used for obtaining the normalised distribution of the FLOB. First, we picked every second text from each genre— a total of 245 texts with approximately 36,750 clauses. Next, we analysed the tense of each clause manually and labelled the clauses as Future and Non-Future; there were 330 instances of Future. We then continued to

identify and count the remaining 36,420 instances to determine which had primary Past tense and which had primary Present tense. Finally, we calculated the ratio of Past tense to Present tense in the LCMC. The normalised distribution was 32:67, which was different from Halliday's (1993:55) result (50:50) in his investigation of the ratio of Past and Present tense in English, and also from the normalised distribution obtained from the FLOB.

Subsequent to calculating the FLOB's normalised tense distribution, the tenses of 535 instances (numbered as DW6) containing 世界 *shi4jie4* in the LCMC were examined and the distribution of the three tenses analysed. There was only one instance (exemplified below) with Past tense, compared to five occurrences of Future tense.

5)	70 年代，	世界	发生
	qi1shi2nian2dai4	shi4jie4	fa1sheng1
	In 1970th,	world	happen
	能源危机	以后，	我国
	neng2yuan2wei1ji1	yi3hou4	wo3guo2
	energy crisis	later	our country
	各炼油厂	进行了	技术改造。
	ge4lian4you2chang3	jin4xing2le1	ji4shu4gai3zao4
	every refinery	conduct	technology reform

In 1970s, every refinery in our country undertook technology reform after confronting the energy crisis.

This strong preference for 世界 *shi4jie4* to be used in Present tense clauses can be

accounted for in two ways. First, compared to Indo-European languages, Chinese is not a tense-obvious language in terms of morphological complexity. This may add to the difficulty of analysing the tense preference or avoidance of 世界 *shi4jie4*. Since the tense was analysed subjectively (according to the sentence particles) by the researcher, the additional difficulty somewhat increased the possibility of error. Second, the major contrast in the use of 世界 *shi4jie4* across the two tenses may be attributable to the greater difference between Past and Present tense in the normalised distribution of Past and Present tense displayed in the LCMC.

However, even taking these factors into consideration, 世界 *shi4jie4* exhibited a much stronger preference for occurring within the use of Present tense than other tenses. Its tense distribution is compared with the normalised tense distribution of the LCMC in Table 6.9 below:

Table 6.9 The Comparison of tense distributions between 世界 *shi4jie4* and normalised distributions of the LCMC

Tense Distribution of 世界 *shi4jie4*		Normalised Distribution of Tense Across LCMC	
Past	5%	Past	32%
Present	94%	Present	67%
Future	0.9	Future	1%

The Table reveals clearly that 世界 *shi4jie4* had a strong preference for occurring with Present tense clauses and a strong aversion to occurring within Past and Future tense clauses. This echoes the data gained from clauses containing *world* in the

FLOB, which revealed *world*'s clear preference for occurring with the same tenses.

6.4 Grammatical functions of 世界 *shi4jie4* in the clause

Using the same methodology as that used to analyse the word *world*, 535 instances of 世界 *shi4jie4* were analysed to see whether they occurred as part of the Subject, as part of the Object, as part of the Complement, or as part of a prepositional phrase functioning as an Adjunct.

Obviously, the percentage of occurrences of 世界 *shi4jie4* in each grammatical position will by itself tell us little about its colligational preferences or aversions unless compared with the grammatical distribution of other nouns. Thus, we selected five lexical items that might operate at a comparable degree of abstractness from the first paragraphs of Text 01 from the LCMC—结果 *jie2guo3* (result), 问题 *wen4ti2* (question), 语言 *yu3yan2* (language), 工业 *gong1ye4* (industry) and 范围 *fan4wei2* (range)—because they were unlikely to have dominant idiomatic uses and or to be used as alternative parts of speech. Although 结果 *jie2guo3* (result) can be occasionally used as a verb, meaning *to kill somebody* or *to fruit blossom*, this kind of use occurs less than five times in the LCMC and therefore will not affect the overall tendency. All instances containing each comparator words were considered, except for 问题 *wen4ti2* (question), 1151 instances of which were found in our corpus. This was even higher than that of the node word 世界 *shi4jie4* and it seemed unnecessary to analyse every instance; thus, we selected half of all instances from each genre, resulting in 451 instances being considered. With regard to the other four

nouns, 242 instances of 结果 *jie2guo3* (result) were considered, 355 instances of 语言 *yu3yan2* (language), 261 instances of 工业 *gong1ye4* (industry), and 141 instances of 范围 *fan4wei2* (range).

Our use of the grammatical categories for Subject, Object, Complement and Adjunct was in line with the normal use defined in Lv Shuxiang's[18] book, 语法修辞讲话 (*Grammar and Rhetoric of Speech Act*). There was shared understanding of the definition of Subject, Object and Adjunct between English and Chinese; however, Complement is more complicated in Chinese than in English. In English, Complement is defined as having the same referent as Subject and typically following the verb BE and other equative verbs such as BECOME. However, the BE group words are only one possibility for manifesting Complement in Chinese. For Chinese, particular adjectives following verbs can also function as Complements, such as 我做**好**了 *wo3zuo4hao3le I have done **well***; in this case, 好 *hao3* (good) is a Complement to the word 做 *zuo4* (do) preceding it.

Another typical characteristic of Chinese is the application of 使动句 *shi3dong4ju4* (officially translated as Causative Sentence, which does not reflect the exact definition of this usage and is not an accurate description of this kind of sentence). From the outset, this sentence follows a typical structure of Subject + Verb + Object; however, the relationship between the Subject and Object in 使动句 *shi3dong4ju4* is not predictive, but emphasises the influence of the Subject on the Object:

---

[18] Lv Shuxiang (吕叔湘), born in 1904, a famous Chinese linguist who devotes all his life into studies of Chinese and dictionary edition. Professor Lv graduated from Oxford University in 1938 and he was also a great linguist in English, who used to teach English at Tsing Hua University at Beijing.

6) 科学仪器　　　　　　　使人的感官　　　　　　得以延伸，

　　ke1xue2yi2qi4　　　shi3ren2degan3guan1　　de2yi3yan2shen1

　　Scientific instruments　people's feelings　　　to be extended

　　使观察者的事业　　　　　　　　　　　　由宏观世界

　　shi3guan1cha2zhe3deshi4ye4　　　　　you2hong2guan1shi4jie4

　　observers' career　　　　　　　　　　from macro-world

　　拓展到　　　　　　　　　　　　　　　宇观世界。

　　tuo4zhan3dao4　　　　　　　　　　yu3guan1shi4jie4

　　extend to　　　　　　　　　　　　　universal world

The scientific instruments extend people's feelings and enable observers' careers to expand from macro-world to the universal world.

In this instance, 感官 *gan3guan1* (feeling) was the Object of 延伸 *yan2shen1* (extend) and 事业 *shi4ye4* (career) was the Object of 拓展 *tuo4zhan3* (expand); however, the Subject and Object in this sentence did not occur in an Subject + Object order, but in a reversed order. The application of this usage was to highlight the influence of a scientific device for studying the universe on a researcher's career by yielding new astronomical findings, in turn benefiting the research. The determination of this kind of case needs further investigation, taking account of the whole sentence or context. Instances belonging to this category were grouped as OTHER due to its low frequency of occurrence.

Results of the grammatical distribution of 世界 *shi4jie4* can be found in Table 6.10, and the comparison of the grammatical distribution of 世界 *shi4jie4* in the clause with that of other nouns is given in Table 6.11. Key results worthy have been

emboldened.

Table 6.10 Results of grammatical distribution of 世界 *shi4jie4* in the LCMC

	Part of Subject	Part of Object	Part of Complement	Part of Adjunct	Other
Percentage	29%	50%	7%	12%	2%
Times of occurrences	155	268	37	64	11

Table 6.11 A comparison of the grammatical distribution of 世界 *shi4jie4* in the clause with that of other nouns in the LCMC

	Part of Subject	Part of Object	Part of Complement	Part of Adjunct	Other
世界 *shi4jie4 world*	14% (75)	**58% (310)**	**7% (37)**	20% (107)	1% (5)
结果 *jie2guo3 result*	36% (88)	23% (55)	9% (22)	18% (44)	14% (33)
问题 *w en4ti2 question*	22% (100)	56% (251)	11% (50)	6% (25)	6% (25)
语言 *y u3yan2 language*	24% (58)	19% (46)	43% (104)	10% (23)	4% (11)
工业 *g ong1ye4 industry*	55% (144)	34% (90)	0	7% (18)	3% (9)
范围 *fan4wei2 range*	17% (23)	40% (56)	0	44% (62)	0

Table 6.11 shows that 世界 *shi4jie4* was quite strikingly different from the other five nouns in terms of grammatical functions. There was a clear positive colligation between 世界 *shi4jie4* and the grammatical function of Object. The second word following 世界 *shi4jie4* occurrence within the Object function was 问题 *wen4ti2* (question), which appeared two percent less frequently than 世界 *shi4jie4*. However, 世界 *shi4jie4* had a relatively negative preference for serving as part of a Complement. Excluding 工业 *gong1ye4* (industry) and 范围 *fan4wei2* (range),

199

which did not occur as part of this function, 世界 *shi4jie4* occurred less than one-tenth of the time as part of a Complement. There was also a negative colligation between 世界 *shi4jie4* and the function of Subject, compared to the other five comparator words. 世界 *shi4jie4* occurred as part of a Subject less than one-fifth of the time, while the other nouns in the sample occurred from one-fifth to more than one-half of the time. To compensate, there was a positive colligation between 世界 *shi4jie4* and the Adjunct function. Only one of the other five nouns — 范围 *fan4wei2* (range) — outweighed the proportion of occurrences of 世界 *shi4jie4* within an Adjunct, and only one noun — 结果 *jie2guo3* (result) — came close to the frequency found for 世界 *shi4jie4*. The others occurred within the Adjunct function three times less often than 世界 *shi4jie4*.

We found positive and negative colligations with the clausal functions with which 世界 *shi4jie4* was likely to be associated; these preferences and aversions are summarised as follows:

There was a clear positive colligation between 世界 *shi4jie4* and the grammatical function of Object, accounting for over half the total instances extracted from the LCMC. There was also a positive colligation between 世界 *shi4jie4* and the grammatical function of Adjunct. Though the percentage was not strikingly higher than for several of the other nouns (excluding 范围 *fan4wei2* (range)), 世界 *shi4jie4* displayed a relatively stronger tendency to be associated with this function.

On the other hand, there was a negative colligation between 世界 *shi4jie4* and the

grammatical function of Complement, especially when compared with 语言 *yu3yan2* (language). A negative colligation was also found between 世界 *shi4jie4* and the grammatical function of Subject; only 14% of instances containing 世界 *shi4jie4* occurred within this function, compared to from 17% to 55% of other nouns.

6.5 Colligational characteristics of 世界 *shi4jie4* within the nominal group

The previous section shows that 世界 *shi4jie4*, like its English equivalent *world*, was colligationally distinct from the comparator nouns in terms of both grammatical preferences and grammatical aversions at the level of clause. This section will look at colligational behaviour of 世界 *shi4jie4* at the rank of the group or phrase. As with *world*, there are, in principle three grammatical possibilities for 世界 *shi4jie4* occurring within a nominal group: as head of the nominal group in which it appears, as part of premodifier, or as part of the postmodification. For example:

7) 当今**世界**   属于   激烈的

dang1jin1shi4jie4   shu3yu2   ji1lie4de

Today's world   belongs to   fierce

市场竞争   当中。 shi4chang3jing4zheng1

dang1zhong1

market competition   in

[世界 *shi4jie4* **occurs as a noun head**]

The world today is involved in fierce market competition

8) 自然神　　　　　成为　　　　　当时　　　　　主宰人们

zi4ran2shen2　cheng2wei2　dang1shi2　zhu3zai3ren2men

The god of nature　　become　　that time　　dominatepeople

内心**世界**的　　　　　　　　主要　　　　精神力量。

nei4xin1shi4jie4de　　　　zhu3yao4　jing1shen2li4liang4

inner world　　　　　　　　major　　　spirit power

[**世界** *shi4jie4* occurs as part of premodifier]

The God of nature had become the main spirit power for people's inner world at that time

9) 上海光源　　　　　　　以**世界**　　　　同类装置

Shang4hai3guang1yuan2　yi3shi4jie4　tong2lei4zhuang1zhi7

Shanghai light source　　world　　　instruments alike

最少的　　　投资　　　和　最快的　　　建设速度，

zui4shao3de　tou2zi1　he2　zui4kuai4de　jian4she4su4du4

least　　　　investment　and　fast　　　construction speed

实现了　　　优异的　　　　　性能…

shi2xian4le　you1yi4de　　　xing4neng2

realise　　　brilliant　　　performance

[**世界** *shi4jie4* occurs as postmodification]

[The company of] Shanghai Light Resource had gained brilliant performance on similar devices in the world with the least investment and fast construction.

This section examines whether Chinese has grammatical functions that were preferred or avoided by the group in which the word or word sequence participated. To answer this, we analysed all the nominal groups within which **世界** *shi4jie4*

appeared in the 535 instances retrieved from the LCMC. To supplement the data, we also incorporated a DW7 database, comprised of 1000 instances extracted from zhTenTen11, for a total of 1535 instances under study. As previously done in the examination of world, the syntactic behaviour of 世界 *shi4jie4* was compared with that of the five comparator nouns used above, namely 结果 *jie2guo3* (result), 问题 *wen4ti2* (question), 语言 *yu3yan2* (language), 工业 *gong1ye4* (industry) and 范围 *fan4wei2* (range). We first compared their grammatical functions in a nominal group, then examined whether the patterns that emerged in the nominal group were likely to be affected by the grammatical function of the noun phrase within the clause.

Table 6.12 A comparison of the grammatical distribution of 世界 *shi4jie4* in the nominal group with that of the other five nouns in the LCMC

	Head of nominal group	Part of the premodification of the nominal group	Part of postmodification of the nominal group
世界 *shi4jie4* world	**33% (508)**	**65% (1002)**	**2% (25)**
结果 *jie2guo3* result	100% (165)	0	0
问题 *wen4ti2* question	75% (338)	19% (85)	6% (28)
语言 *yu3yan2* language	53% (127)	43% (115)	0
工业 *gong1ye4* industry	46% (120)	54% (141)	0
范围 *fan4wei2* range	92% (129)	8% (12)	0

Table 6.12 shows that 世界 *shi4jie4* was clearly different from the other nouns under study in terms of its distribution in a nominal group, and that those differences were much more significant and striking than with regard to other grammatical functions. First, while the other nouns all occurred most frequently as heads in nominal groups, particularly 结果 *jie2guo3* (result), 世界 *shi4jie4* occurring as head much less often,

indicating 世界 *shi4jie4* tends to be used to narrow down noun heads and tends to avoid being the centre of attention. Second, 世界 *shi4jie4* had a very strong tendency to occur in the premodification function of a noun group, closely followed by 工业 *gon1ye4* (industry). Third, none of the nouns under study showed a preference for occurring as part of a postmodification, except for 世界 *shi4jie4* and 问题 *wen4ti2* (question), which showed some small tendency to occur in this grammatical position. In summary, 世界 *shi4jie4* had a negative colligation for occurring as noun head, and 世界 shi4jie4a positive colligation for occurring as part of premodification.

6.6 Grammatical distribution of 世界 *shi4jie4* in the nominal group when serving different grammatical functions

In Section 6.5, we examined whether 世界 *shi4jie4* occurred with higher or lower frequency than normal in any of the three positions—i.e., head, part of premodifier, or part of postmodification. The data has shown a colligational association between 世界 *shi4jie4* and these three positions. This section looks at the distribution of these three positions across the grammatical functions of Subject, Object, Complement and Adjunct to investigate whether 世界 *shi4jie4* had a preference for or aversion to occurring at each position while serving different grammatical functions. We excluded instances containing 结果 *jie2guo3* (result) from the comparison due to its extremely significant preference for occurring as a head in a nominal group; in our sample, in all instances containing 结果 *jie2guo3* (result) it appeared as a noun head, making exploring its grammatical position within each grammatical function uninteresting.

Comparison of the remaining four comparators using the 535 instances of 世界 *shi4jie4* revealed there were colligational preferences and aversions for 世界 *shi4jie4* occurring at different positions across different grammatical functions, as displayed in Table 6.13 below.

Table 6.13 The comparison of 世界 *shi4jie4* occurring at different positions in a nominal group while serving as different grammatical functions

	世界 *shi4jie4* world	问题 *wen4ti2* question	语言 *yu3yan2* language	工业 *gong1ye4* industry	范围 *fan4wei2* range
Subject	(155)	(100)	(58)	(144)	(155)
head	47% (32)	75% (75)	60% (35)	50% (72)	100% (23)
part of premodification	41% (63)	0	40% (23)	50% (72)	0
part of postmodification	13% (19)	25% (25)	0	0	0
Object	(268)	(251)	(46)	(90)	(268)
head	24% (64)	67% (167)	67% (31)	50% (45)	79% (44)
part of premodification	75% (200)	33% (84)	33% (15)	50% (45)	21% (12)
part of postmodification	2% (4)	0	0	0	0
Complement	(37)	(50)	(104)	(0)	(37)
head	13% (5)	100% (50)	63% (65)	0	0
part of premodification	87% (32)	0	37% (39)	0	0
part of postmodification	0	0	0	0	0
Adjunct	(107)	(25)	(23)	(18)	(107)
head	45% (49)	100% (25)	0	0	100% (62)
part of premodification	55% (58)	0	100% (23)	100% (18)	0
part of postmodification	0	0	0	0	0

世界 *shi4jie4* had a greater aversion to functioning as either Subject or Object in the form of a noun head than any of the other five words. However, when occurring as part of a premodification in the nominal group, 世界 *shi4jie4* preferred to occurred within the grammatical function of Complement. The findings of this section and the results from Section 6.2.3 can be combined and summarised as follows:

1. 世界 *shi4jie4* had a greater negative colligation for occurring as a noun head than the other five nouns, and an aversion to functioning as Subject and Object when appearing as a noun head.

2. 世界 *shi4jie4* had a positive colligation for occurring as a premodification and a preference for functioning as Complement when doing so. Only one of the other grammatical functions — Object — came close to it in terms of proportion.

This research result can also be interpreted as, when functioning as a part of a Subject or Object, 世界 *shi4jie4* had an aversion to appearing in the position of a noun head in the nominal group. When functioning as a part of an Object, it had a stronger preference for occurring as part of a premodification in the nominal group. This preference for occurring as part of a premodification also applied when 世界 *shi4jie4* served as part of a Complement. This preference was far stronger than those of the other nouns when occurring as part of the same grammatical function. 世界 *shi4jie4* preferred to occurred in the form of a premodification when serving as part of an Adjunct than when occurring in the form of a noun head. The following figure depicts the colligational preferences and avoidances of 世界 *shi4jie4* across grammatical functions:

	Premodification	Noun head
Subject	×	—
Object	+	—
Complement	+	×
Adjunct	+	—

where "×" refers to no significant preference/avoidance occurring, "+" indicates a positive colligation for a particular grammatical function and "－" refers to a negative colligation.

In the process of priming a word or word sequence, individuals may also prime grammatical constraints. As a word or word sequence may be primed to co-occurred with another word/or sequence, it may be primed to occur in or with a particular grammatical function. The major feature of the colligation as defined in LPT is that it goes beyond traditional grammatical relations to embrace the preference/avoidance of a word/word sequence's positioning within a sentence; i.e., colligational statements can be negative and positive.

Our data confirms partially the features of colligation stated in the theory, in that 世界 *shi4jie4* manifested a preference/avoidance for given grammatical positions. When used within grammatical functions, it also exhibited a preference/avoidance for serving as part of a particular function. The remainder of this chapter seeks to show how a colligational description might proceed for a word sequence.

6.7 Colligational behaviour of 世界上 *shi4jie4shang4* (in the world) and *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world)

In Section 6.3, we investigated these two nestings from the perspective of their collocational behaviour. Like the prepositional phrases *in the world* and *around the world*, the Chinese equivalences also had a high frequency of occurrence among all the word combinations containing the node word 世界 *shi4jie4*. We chose these two-word sequences for two reasons. First, both 世界上 *shi4jie4shang4* (in the

world) and *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) ranked among the top-five most-frequently used clusters containing *世界 shi4jie4* in the Chinese corpus of zhTenTen11. Second, these two combinations were recurrent translational versions of their English equivalents *in the world,* with a high MC value — 32% and 49%, respectively.

As was done previously, two databases—DW8 and DW9—were created for the study, each of which contained 1,000 instances retrieved from zhTenTen11. Five to eight instances were extracted from each concordance page showing *世界上 shi4jie4shang4 in the world* and *世界范围内 shi4jie4fan4wei2nei4* (in the range of the world), from different sources at each page across genres.

First, the sentence-positions in which these two combinations liked to appear were examined to determine whether a word sequence had a preferred (or avoided) position in a sentence. Second, we explored whether a word sequence participated in a particular grammatical function and its preference for (or avoidance of) this function. Thirdly, the grammatical position in a nominal group was analysed to examine whether there were a positive (or negative) colligational association between the word sequence and a particular grammatical position.

6.7.1 Colligational behaviour of *世界上 shi4jie4shang4* (in the world)

One thousand of 6,374 instances were examined to identify the potential colligations of *世界上 shi4jie4shang4* (in the world). Fifty-two percent (520 out of 1000

instances) of the instances occurred within the grammatical function Complement, which differed greatly from the overall trend of 世界 *shi4jie4* occurring in this grammatical function. We saw previously that there was a negative colligation between 世界 *shi4jie4* and the grammatical function Complement; however, this aversion became a preference when 世界 *shi4jie4* collocated with 上 *shang4* (up) to form a word sequence.

Connecting with this colligational preference was a positional feature: 世界上 *shi4jie4shang4* (in the world) apparently preferred to occur in the middle of a clause, especially after the typical words (e.g., 是 *shi4 Be*) marking a Complement (336 out of 520), for example:

10) 在各个国家中，	印尼	是	**世界上**
zai4ge4ge4guo2jia1zhong1	yin4ni2	shi4	shi4jie4shang4
Among every country	Indonesia	was	in the world
火上	最多的		国家，
huo3shan1	zui4duo1de		guo2jia1
volcano	most		country
现有 129 座	活火山。		
xian4you3129zuo4	huo2huo3shan1		
now had 129	active volcanoes		

Out of all the countries in the world, Indonesia is the country which has the most volcanoes. Currently, it has 129 active volcanoes.

When occurring in a Complement, 世界上 *shi4jie4shang4* (in the world) had a

strong preference for being followed by the word 最 *zui4* (a typical superlative marker in Chinese). The superlative meaning depended on the adjective subsequent to the word 最 *zui4*, such as 最大 *zui4da4* (the biggest) or 最好 *zui4hao3* (the best). These superlatives, together with words such as 第一 *di4yi1* (first), belonged to a semantic set forming the semantic association of EVALUATION.

There was also a negative colligation between 世界上 *shi4jie4shang4* (in the world) and the Subject and Adjunct grammatical functions. The proportions of 世界上 *shi4jie4shang4* (in the world) occurring in these two functions were coincidentally the same — 13% (130 out of 1000 instances).

As far as grammatical positions in the nominal group were concerned, 74% (740 out of 1000 instances) were found to occur as part of a premodification, which overwhelmed any other grammatical positioning. With all the colligational features of 世界上 *shi4jie4shang* examined, its colligations were then:

● There was a negative colligation between 世界上 *shi4jie4shang* and the Subject and Adjunct grammatical functions, which was in line with that of 世界 *shi4jie4* occurring as an individual word

● To compensate, there was a clear positive colligation between 世界上 *shi4jie4shang* and the Complement function that was four times stronger than that of the other grammatical functions

● 世界上 *shi4jie4shang4* (in the world) colligated positively with premodification, while colligating negatively with postmodification (no instance) and head (only

10 out of 1000 instances occurred in this grammatical position)

- 世界上 *shi4jie4shang4* (in the world) had a strong textual colligation with middle position in a clause

- 世界上 *shi4jie4shang4* (in the world) had an aversion to being pre- or post-modified in any position

6.7.2. Colligational behaviour of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world)

There were 13,591 instances containing *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) (6.45 per million running words) found in the Chinese corpus of zhTenTen11. As the number of concordance lines containing this word combination was unmanageable within the thesis's scope, a database containing 1,000 extracted instances from the corpus was created for detailed analysis, as in previous sections and chapters. This database was subsequently named DW10.

Regarding the grammatical functions it served and its sentence positions, the colligational picture of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) was very different from that of 世界上 *shi4jie4shang4* (in the world). While 世界上 *shi4jie4shang4* (in the world) had a negative association with the Adjunct grammatical function, 48% of the instances of *（在）世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) showed a positive colligation with this function. The ratio between *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in

the range of the world) and 世界上 *shi4jie4shang4* (in the world) serving as Adjunct was around 4:1.

Another difference was that the colligation of *（在）世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) within the grammatical function Complement was strikingly reversed, with only 2% (20 out of 1000 instances) serving this function. In sharp contrast, 世界上 *shi4jie4shang4* (in the world) occurred most frequently (72%) in this grammatical function as part of a premodification than it did within other functions—36 times more than that of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) occurring in a Complement. While having a negative colligation with occurring within the Complement grammatical function, the combination of *（在）世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) colligated positively with the Adjunct grammatical function. This colligational preference corresponds to the colligational preference (shown in Table 6.11) of 范围 *fan4wei2* (range)—a component making up the word combination studied in this section. Table 6.11 also reveals that 世界 *shi4jie4*, as a key element of this combination, comes closest to 范围 *fan4wei2* (range) in terms of its association with the Adjunct grammatical function. This may explain the high frequency of the combination including these two individual words associated with the Adjunct grammatical function.

Connected with its colligational preference for occurring within the Adjunct grammatical function, *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) was also likely to appear at the initial position of a clause, for example,

10) 在   世界   范围内，   曾   发生

zai4   shi4jie4   fan4wei2nei4   ceng2   fa1sheng1

In   world   range   had   happened

两起   核事故。

liang3qi3   he2shi4gu4

two   nuclear accident

Two nuclear accidents have happened in the world.

Eighty-four percent of the instances in which it was used as an Adjunct occurred at the very beginning of a clause. Of those instances, 202 (42%) occurred in the form of 从...来看 *cong2...lai2kan4* (seeing from…), as exemplified in instance 11:

11) 从   世界   范围   来看，

cong2   shi4jie4   fan4wei2   lai2kan4

From   world   range   to see

以研究人类   为   对象的   人类学

to study human race   as   Object   anthropology

yi3yan2jiu1ren2lei4   wei2   dui4xiang4de   ren2lei4xue2

在本世纪 30 年代   就开始   进入城市。

zai4ben3shi4jie430nian2dai4   jiu4kai1shi3 jin4ru4cheng2shi4

at the beginning of 1930s   start to   enter into cities

The world anthropology which studies the human race had taken urban areas as the starting point at the beginning of 1930s.

The remaining instances of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range

of the world) used in the Adjunct function (125 out of 480 instances, 26%) took the middle position of a sentence, as illustrated in below:

12) 到了　　　　　　　19 世纪末　　　　　　20 世纪初，

dao4le1　　　　　　　19shi4jie4mo4　　　　　20shi4jie4chu1

To　　　　　　　　　the end of 1800s　　　　the beginning of 1900

西欧　　　　　　　　殖民体系　　　已在　　**世界范围内**

xi1ou1　　　　　　　zhi2min2ti3xi4　yi3zai4　shi4jie4fan4wei2nei4

Western European　　colonial system　had　　in the range of the world

建立，　　从而　　　完成了　　　　资本主义　　　向

jian4li4,　　cong2er3　wan2cheng2le　zi1ben3zhu3yi4　xiang4

establish,　　so as　　complete　　　capitalism　　　to

全球　　　　　　　　扩张的　　　　　　　　　　过程。

quan2qiu2　　　　　　kuo4zhang1de　　　　　　　guo4cheng2

the world　　　　　　expand　　　　　　　　　　process

　　At the end of 1800s and the beginning of 1900s, Western Europe had established a world-wide colonial system, thus completing the global outspread process.

All instances of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) within grammatical functions of Subject and Object appeared as part of a premodification. Examples of the combination occurring within the two grammatical functions, respectively, are illustrated below:

13) 当前，　　　　　**世界范围**的　　　　　经济科技竞争

dang1qian2　　　shi4jie4fan4wei2de　　　jing1ji4ke1ji4jing4zheng1

At present　　　in the world　　　　　　economy and science

异常	激烈。
yi4chang2	ji1lie4
extremely	fierce

Currently, the competition of economy and science is extremely fierce in the world.

14) 对 可再生 能源的

dui4	ke3zai4sheng1	neng2yuan2de
For	hydropower	resources
研究	和	利用
studies	and	applications
yan2jiu1	he2	li4yong4
受到	**世界范围**的	重视。
shou4dao4	shi4jie4fan4wei2de	zhong4shi4
gain	world-wide	attention

Studies and applications for hydropower have gained a world-wide attention.

Based on the above research results, the following were the typical colligational features of *(在) 世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world):

● a strong colligation with an Adjunct;

● a strong aversion to occurring with a Complement;

● a strong association with first position in the sentence (and the clause) when serving as an Adjunct;

● an aversion to being postmodified in any position of a clause; and,

● a strong tendency to premodify the noun head when part of a Subject and Object.

We have shown previously, using the data found for 世界 *shi4jie4,* that there was a positive/negative colligation for this node word occurring within a particular grammatical function at different positions in a nominal group. The data in this section supplement those results, in that we have shown combinations containing 世界 *shi4jie4* also had a preference/avoidance for a particular position in a sentence or clause. The colligational behaviour of a word sequence might also be affected by the colligational behaviour of the components from which it is made. It is apparent, based on our consideration of 世界 *shi4jie4,* that nesting, as defined in LPT, can not only take the form of a combination built out of word sequences, but also that of the colligational behaviour of individual words. This phenomenon is accounted for as colligational nesting in LPT, which claims that "a word or word sequence combines with a particular colligational priming (positive or negative), this nesting in turn had further primings, which may be of any kind— collocational, semantic associational or colligational" (Hoey 2005: 58). If we apply this to account for the results of this study, it would appear that the nesting of 世界 *shi4jie4* within a Complement was primed to collocate with the word 是 *shi4 BE* ahead of it and 最 *zui4* following it, and that this nesting was further primed to collocate with a semantic set forming a semantic association to EVALUATION.

Given that we have seen that *world* had a variable distribution across each grammatical function and that there were collocational and semantic associational choices dependent on which function was chosen, it makes sense to investigate whether nestings containing 世界 *shi4jie4* behaved differently regarding the

216

collocations and semantic associations chosen for a particular grammatical function.

The next chapter considers the occurrence of 世界 *shi4jie4* in complex situations and investigates its semantic associations by combining its colligational behaviour. Its textual semantic associations and nestings will also be examined.

CHAPTER 7

世界 *shi4jie4* in Complex Situations

7.1 Introduction

In Chapter 6, we provided a general collocational and colligational profile of 世界 *shi4jie4* in terms of its left and right side within a five-word span. It was noted that there were a few collocates and colligations shared with English at the L1, L2 and R1, R2 positions, even though Chinese was normally considered a markedly different language from English in many respects. As with *world*, 世界 *shi4jie4* (6.681) collocated with the preposition 在 *zai4* at the L2 position and with 最 *zui4* (7.602) at the R2 position.

To obtain fuller data about 世界 *shi4jie4* in combination with its collocational and semantic associational features and to uncover data to compare with that of *world,* examined in Chapters 4 and 5, the following sections investigate 世界 *shi4jie4* within more complex situations.

7.2 The semantic associational profile of 世界 *shi4jie4*

Tables 6.1-6.3 shows that the left collocates to 世界 *shi4jie4* were 全 *quan2* (all)*,* 当今 *dang1jin1* (today)*,* 成为 *cheng2wei2* (become), etc.*;* Tables 6.4-6.6 show that the right collocates to 世界 *shi4jie4* were 一流 *yi1liu2* (first level)*,* 最 *zui4* (most)*,* 上

shang4 (up), 国 *guo2* (country), 遗产 *yi2chan3* (heritage), and so forth. Collocates appearing at each position were then available to form semantic sets which could form particular semantic associations in turn.

The first and largest of the semantic associations of 世界 *shi4jie4* at the L1 position identified in zhTenTen11 was a class of adjectives that located 世界 *shi4jie4* in different TIME PERIODS. This semantic association comprised 64% of all premodifying adjectives, including collocates with high LogDice scores such as 当今 *dang1jin1* (today) (8.679), 目前 *mu4qian2* (at present) (6.919), 当代 *dang1dai4* (contemporary) (6.344) and collocates with lower LogDice scores such as 当前 *dang1qian2* (currently) and 当时 *dang1shi2* (at that time) (4.996):

1) 公司　　　　引进了　　**当今世界**　　　　最　　专业的

   gong1si1　　yin3jin4le1　**dang1jin1shi4jie**　zui4　zhuan1ye4de

   The company　introduce　**today's world**　　most　professional

   精密仪器　　　　和　　　　辅助材料。

   jing1mi4yi2qi4　　he2　　fu3zhu4cai2liao4

   accurate instruments　and　　auxiliary materials.

   The company has introduced the most professional and accurate instruments in today's world.

2) **当代世界**　　　　各国　　　面临着　　　诸多

   dang1dai4shi4jie4　ge4guo2　mian4lin2zh3　zhu1duo1

   Contemporary world　every country　face　　　　many

   要靠　　　共同协作　　　　　　才能　　应付的

   yao4kao4　gong4tong2xie2zuo4　cai2neng2　ying4fu4de

yao4kao4   gong4tong2xie2zuo4   cai2neng2   ying4fu4de

need   collaborative cooperation   to   deal with

问题…

wen4ti2

problem

Countries in the contemporary world are facing many problems that may need collaborative cooperation to deal with.

The phenomenon of both higher- and lower-score collocates forming the same semantic set and being generalised into the same semantic association confirms Hoey's (2005: 17) claim that when we formulate what we want to say, primings shape the wording we use. Collocation can account for the routine of speakers' choices for a word sequence; however, semantic association can account for some aspects of creativity. The circular of meaning and word choices can be accounted for with priming, in that collocation forms an individual's (or speakers in the same community) semantic association, which in turn is primed while the speaker intends to communicate in a particular context. Because speakers have different communicative needs or linguistic experiences, they may prime words which are not used frequently by others but belong to the semantic association made up of the routine words to realise their communicative goals. There are co-occurrences which cannot be accounted for in terms of collocation in Chinese. Semantic association is a necessary generalisation that appears to reflect Chinese speakers' lexical priming.

The second semantic association suggested by the data was that 世界 *shi4jie4* associated with adjectives expressing a person's MIND. This category accounted for 23% of the adjectives examined, including items such as 内心 *nei4xin1* (inner) and

灵魂 ling2hun2 (soul).

Example indicating this kind of semantic association was shown as follows:

3) 我们　　　可以　　　通过　　　参与者的　　　对话，

   wo3men　ke3yi3　tong1guo4　can3jia1zhe3de　dui4hua4

   We　　　could　　by　　　participants'　conversation

   了解　　　　　参与者的　　　　**内心世界**。

   liao3jie3　　can1yu3zhe3de　**nei4xin1shi4jie4**

   understand　participants　　**inner world.**

We could understand the participants' inner world through their conversations.

The third largest semantic association referred to a COUNTRY, REGION or an INTERNATIONAL ORGANISATION. This semantic association accounted for 11% of the data and was formed by collocates such as *西方的 xi1fang1de* (Western), *阿拉伯 a1la1bo2* (Arabian), and *联合国 lian2he2guo2* (United Nations). Examples include:

4) **西方世界**　　　对　　　大熊猫的　认识

   Western world　to　　Panda　　knowledge

   始于　　　　　四川宝兴。

   starts from　　Sichuan　　Baoxing

Western world's knowledge of Panda originates from Baoxing, Sichuan province.

5) 长期以来，　　　中国的针灸术一直受到　　　　**西方世界的**

   In the long run,　Chinese acupuncture　　　has been　**Western world**

怀疑。

doubt.

In the long run, Chinese acupuncture has been doubted by Western world.

7.3 Textual semantic association of 世界 *shi4jie4*

Chapter 5 showed the combinations *world* + SHRINK，*world* + OWE and *world* +

KNOW had textual semantic associations in a more distant neighbourhood. The

combination *world + OWE* seemed to associate with a textual association with

emphasis; for example, *The world doesn't owe you. You should owe the world.*

Although textual semantic association was not referred to in the title of this study, we

should not be in a rush to dismiss discourse issues as independent of and

unconnected to what we have discovered in the study.

Having found evidence to support Hoey's argument on textual semantic association,

we then asked, *Does this feature also apply to Chinese?* To answer this, we

investigated instances belonging to the three semantic associations and retrieved

instances with significant combinations containing collocates with a semantic

association with TIME PERIOD. In total we found *23* instances of 当今世界

dang1jin1shi4jie4 (today's world), 13 instances of 当代世界 *dang1dai2shi4jie4*

(contemporary world), 18 instances of 目前世界 *mu4qian2shi4jie4* (present world)

and five instances of 当时世界 *dang1shi2shi4jie4* (world at that time) extracted from

the LCMC. This database was then numbered as DW10.

Our mini-corpus showed that the first two combinations containing 世界 *shi4jie4*
were dominantly associated with a textual meaning of COMPARISON at a span
wider than +/-5, accounting for 97% and 83% of this discourse function,
respectively:

6) 这不是　　　　什么　　　　畅想，　　　　　而是　　　　**当今世界**

   zhe4bu2shi4　　shen2me　　chang4xiang3　　er3shi4　dang1jin1shi4jie4

   It was not　　　　　what　　　　imagination,　　but　　**today's world**

   已付诸　　　　实践的　　　　事实。

   yi3fu4zhu2　　shi2jian4de　　shi4shi2

   has put into　　　practice　　reality.

It was not an imagination, but a reality that the current world has put it into practice.

7) 事实上，　　　　　由于　　　**当代世界**　　　　　与

   shi4shi2shang4　　you2yu2　　dang1dai4shi4jie4　　yu3

   As a matter of fact,　because　**contemporary world**　and

   二次世界大战前　　　　　世界之间的差别　　　　　　要比

   er4ci4shi4jie4da4zhan4qian2　　shi4jie4zhi1jian1cha1bie2　　yao4bi3

   before the Second World War　　world differences　　　　more

   当时　　　　美洲与欧洲的差别　　　　　　　　大得多…

   dang1shi2　　mei3zhou1yu3ou1zhou1decha1bie2　　da4de3duo1…

   at that time　　America and Europe difference　　bigger…

As a matter of fact, the differences occurring in the contemporary world was much
bigger than differences between America and Europe before the second world war.

The other two combinations were typically associated with a meaning of

EVALUATION, with percentages of almost 100% and 70% respectively:

8) 它是　　　目前世界　　　最大的　　　核电站。

ta1shi4　　mu4qian2shi4jie4　zui4da4de　　he2dian4zhan4

It was　　　**present world**　largest　　　power station.

It was the largest power station in the present world.

9) 到　　西周　　　　　　　　　　　　鼎盛时期

dao4　xi1zhou1

ding3sheng4shi2qi1

To　　Xizhou Dynasty (1100 B.C. to 771 B.C.)　　peak time

中国人口　　　　　　　已达　　　　　　　1300 万左右，

zhong1guo2ren2kou3　yi3da2

1300wan4zuo3you4

Chinese population　　　has reached to　　　　about 130 million,

是　　　当时世界上　　　　　人口最多、疆域最大的

shi4　　dang1shi4shi4jie4shang4　ren2kou3zui4duo1

jiang1yu4zui4da4

was　　　**world at that time**　the largest in population and

territory

文明国家。

wen2ming2guo2jia1

civilised country.

To the peak time of Xizhou Dynasty (1100 B.C. to 771 B.C.), Chinese population has reached to about 130 million, which makes it the civilised country with the largest population and territory.

The second largest semantic association comprised semantic sets of adjectives associated with human beings' MIND. The representative collocates were 内心 *nei4xin1* (inner) and 心灵 *xin1ling2* (mind). Similar to the establishment procedure for DW3, we created a database (subsequently named DW11) to examine the textual semantic association with 内心世界 *nei3xin1shi4jie4* (inner world) and 心灵世界 *xin1ling2shi4jie* (mind world). Seven examples of the first combination and five of the second were retrieved from the LCMC. Our data showed these two combinations were associated with a textual meaning of STUDY or REVEAL a FIGURE (both having a prototype in reality or created by a writer) in a literary work or a group of people (e.g. poets). Examples illustrating the first possibility are:

10) 既要　　分析　　　　人物　　　**内心世界**　　　　　复杂活动，

　　ji4yao4 fen1xi1　　ren2wu4　nei4xin1shi4jie4　　fu4za2huo2dong4

　　both　　analyse　　character's　**inner world**　　complicated activity

　　又要　　　解释　　　　　　　人物与外部世界

　　you4yao4　jie3shi4　　　　　ren2wu4yu3wau4bu4shi4jie4

　　and　　　explain　　　　　　character and outside world

　　种种联系。

　　zhong3zhong3lian2xi4

　　every kind of relationship

[The character here refers to Queen Wu Zetian (624-705 A.D.), the most influential female empire in Chinese history].

We need to analyse the complicated inner world of the character as well as to explain the relationship between the character and the environment she was living in.

An example illustrating the second case is:

11) 它　　　主要　　　解释　　　作者的　　　**心灵世界 —**

ta　　　zhu3yao4　　jie4shi4　　zuo4zhe3de　　xin1ling2shi4jie4

It　　　mainly　　explains　　writer's　　　**mind world** —

内心的空虚　　　　　　所引起的　　　噩梦般的恐惧。

nei4xin1dekong1xu1　　suo3yin3qi3de　　e4meng4ban1dekong3ju4

emptiness of mind　　　caused by　　　nightmare-like fear

["It" here refers to the autobiographical book]

It mainly explains the writer's mind world — a nightmare-like fear caused by the emptiness of mind.

The third and final semantic association with premodifiying adjectives was made up of words indicating a COUNTRY, REGION or ORGANISATION. A group of typical collocates was *西方 xi1fang1* (Western)*, 阿拉伯 a1la1bo2* (Arab) *and 联合国 lian2he2guo2* (United Nations); 23 instances of *西方世界 xi1fang1shi4jie4* (Western world), eight instances of *阿拉伯国家 a1la1bo2guo2jia1* (Arabian country), and 32 instances of *联合国世界 lian2he2guo2shi4jie4* (UN world) were extracted, from which the W5 was established. There were two textual semantic associations expressed by a further distance of *西方世界 xi1fang1shi4jie4* (Western world), namely CULTURAL COMMUNICATION (five out of 23) and SHOWING A POLITICAL POINT OF VIEW[19] (nine out of 23), and one textual semantic association with COMPARISON (11 out of 23).

Examples illustrating each textual semantic association are displayed below in

---

19 The political views expressed in the examples are those of the speaker only, and do not represent the views of the author.

examples 12 to 14:

12) 当整个**西方世界**　　　　　　正在急速发展的时候，

dang1zheng3ge4xi1fang1shi4jie4　zheng4zai4ji2su4fa1zhan3de1shi2hou4

When the whole **Western world**　　was rapidly developing,

我们的国家　　　　　　　　却在　　　　　　《马关条约》的

wo3men2deguo2jia1　　　　que4zai4　　　　ma3guan1tiao2yue1de

　　our country　　　　　　however　　　　Shimonoseki Treaty

　　割地赔款中　　　　　　　苟延残喘。

　　ge1di4pei2kuan3zhong1　gou3yan2can2chuan3

　　ceding territory and extortion float dimly.

　[Shimonoseki Treaty was a Sino-Japan treaty that required China to cede

　Taiwan to Japan]

When the whole Western world was exerting all the efforts on the development,
China, however, floated dimly in the consequence caused by the Shimonoseki Treaty.

13) 中国　　　许多　　　瓷器　　　被传播到**西方世界**，

zhong1guo2xu3duo1　　ci2qi4　　bei4chuan2bo1dao4xi1fang1shi4jie4

Chinese　　many　　　porcelains　were spread to **Western world**

不仅如此，　　　　　还有　　　　我们的　　　儒家文化。

bu4jin3ru2ci3　　hai2you3　　　wo3men2de　　ru2jia1wen2hua4

not only this,　　and also　　　our　　　　Confucian culture

Not only has chinaware spread to Western world, but also our Confucian culture.

14) 中国的　　　崛起　　　　引起　　　**西方世界**的 恐慌。

zhong1guo2de　jue2qi3　　yin3qi3　　xi1fang1shi4jie4dekong3huang1

China's development aroused **Western world** panic

China's development has aroused the panic of Western world.

The combination *阿拉伯世界 a1la1bo2shi4jie4* (Arabian country) had a textual semantic association with CALLING for SOLUTION. All examples derived from our data (eight instances) show that speakers using this combination tended to show an attitude of calling for leaders of every country and, in particular, of Arabian countries (especially those with religious conflicts) to seek solutions to conflicts or problems in the Arab world:

15) 对于 **阿拉伯世界** 所面临的问题，

dui4yu2 a1la1bo2shi4jie4 suo3mian4lin2de1wen4ti2

For **Arabian world** facing problems,

各国领导人 在求同存异的

ge4guo2ling3dao3ren2 zai4qiu2tong2cun2yi4de

leaders from each country seek common ground while reserving

differences

基础上 统一了看法。

ji2chu3shang4 tong3yi1le1kan4fa3

on the basis of unite opinions

Leaders from each country have united their opinions on the problems and conflicts of Arabian world based on the principle of "seeking common ground while reserving differences."

An examination of the third typical combination containing collocates associated with a meaning of REGION (e.g. 联合国世界 lian2he2guo2shi4jie4 (UN world)) showed it had a semantic association with EVALUATION at a greater distance. The

textual semantic association was categorised into two sub-classes, one emphasising the great achievements of an organisation and one showing the authority of an organisation:

16) 中国书法　　　　　　被列为　　　　　**联合国世界**文化遗产。

zhong1guo2shu1fa3　　bei4lie4wei2　　lian2he2guo2shi4jie4wen2hua4 yi2chan3

Chinese calligraphy　　was listed as　　**UN world** culture heritage

Chinese calligraphy was declared a World Culture Heritage by the United Nations.

17) 根据　　　　　　　　　　**联合国世界**旅游组织预测，

gen1ju4　　　　　　　　lian2he2guo2shi4jie4lv3you2zu3zhi1yu4ce4

According to　　　　　**UN World** Tourism Organisation Prediction

中国的旅游业　　　　　　将会进一步发展。

zhong1guo2delv3you2ye4　　jiang1hui4jin4yi2bu4fa1zhan3

China's tourism industry　　will develop further

The UN World Tourism Organisation predicted that China's tourism industry would develop further.

The textual semantic association feature detected from 世界 *shi4jie4* in this section was quite unlike the local semantic association examined in Chapter 6. The long-distance relationship showed a typical interaction between the reader and the writer. Examination of this issue suggested a priming between lexis and text.

Chinese has long been argued a paratactic language, due to its rather loose connection to grammatical relationships. On the contrary, English is normally seen as

a hypotaxis language that is systematically and tightly connected by its grammatical relationships. A number of studies contrasting English and Chinese have detected how Chinese parataxis operates differently from English hypotaxis. However, none has explained why Chinese tends to be connected by semantic meaning. The discovery of textual semantic association for 世界 *shi4jie4*, however, may suggest LPT's ability to account for the parataxis of Chinese. The evidence found for 世界 *shi4jie4* shows this lexical item (or combinations made from this item) can be positively or negatively primed to occurred as part of a specify type of semantic or pragmatic relationship (while pragmatic relationships are not the focus of this study, the boundary between semantic and pragmatic is rather blurry. Here we address both to refer to the potentiality of pragmatic relationships) or in a specific textual pattern. Textual semantic associations may occurred between clauses or between larger chunks of text. Compared with English, whose semantic relations is realised within a shorter distance, semantic relationships in Chinese (as evidenced by 世界 *shi4jie4*) may be realised at greater distances (e.g. 15 tokens). This more distant connection of lexis may move linguists to conclude that Chinese semantic meaning is linked with loose systematic formality. Indeed, the positive and negative collocation and colligation detected from 世界 *shi4jie4* show that systematic connections in Chinese may not be as loose as thought; they just may be a bit farther from the word in focus.

7.4 Collocational profile of 世界 *shi4jie4* when occurring within Subject

Chapter 6 showed the distribution of 世界 *shi4jie4* at different grammatical positions across different functions. We discovered that 世界 *shi4jie4* had a negative preference

for occurring within the Subject and a Complement functions and a positive preference for occurring within the an Object and Adjunct functions. Within each grammatical function, 世界 *shi4jie4* occurred more dominantly at the position of premodification than the comparator nouns. Its greatest preference for appearing at this position was within a Complement ($p<0.05$).

This and subsequent sections detail the collocational and semantic associational behaviour of 世界 *shi4jie4* appearing at different grammatical positions while serving as a Subject.

7.4.1 Collocations and semantic associations when occurring as noun head

Our sample corpus (made up with instances extracted from the LCMC) showed only 16 out of 75 instances containing 世界 *shi4jie4* at the noun head position within the Subject function. 世界 *shi4jie4* in all these instances co-occurred with the verb 变得 *bian4de4* (has become). To obtain more cases, we used Sketch Engine to retrieve concordance lines for our analysis. zheTenTen showed 1,050 (LogDice score 6.85) instances containing 世界 *shi4jie4* occurred within the Subject function at the noun head position. Sketch Engine also revealed that 715 instances containing 世界 *shi4jie4* appeared as a noun head within a Subject, followed by the word 变成 *bian4cheng2* (change into/transform). The remaining instances of 世界 *shi4jie4* appearing at the noun head position were followed by the verb 变得 *bian4de3* (has become). These two dominant words were grouped into a semantic set associated

with CHANGE. Of all the instances containing 世界 *shi4jie4* + 变得 *bian4de* and 世界 *shi4jie4* + 变成 *bian4cheng2* within the Subject, 24% occurred with the definiteness marker 这个 *zhe4ge4* (this).

7.4.2 Collocation and semantic association when occurring as premodification

67% (50 out of 75) instances containing 世界 *shi4jie4* were part of a premodification within a Subject. Seventy-four percent (37 out of 50) occurred as a noun modifier within the nominal group of premodification, while the remaining instances occurred in the form of 世界(上)的... *in the world*. Interestingly, the words collocating with 世界 *shi4jie4* in the nominal group of premodification were all verbs, such as 领先 *ling3xian2* (to pioneer); 首创 *shou3chuang4* (to create something first time in the world), 惊叹 *jing1tan4* (to amaze) and 首屈一指 *shou3u1yi4zhi3* (to be the first and foremost). These words were all associated with a meaning of HIGH RANKING.

Unlike English, which makes great use of postmodification, Chinese is unlikely to use it. Among the few cases, only nine instances of 世界 *shi4jie4* appearing at this position within a Subject were discovered. The postmodification occurred in the structure of 以 *yi3*…, for example 他以世界第一的身份出席比赛 *He was present at the competition as the "world's best" player*. From the nine instances containing 世界 *shi4jie4* as a postmodification, 世界 *shi4jie4* appeared as a noun modifier of the nominal group, positioning as part of a postmodification.

The first thing to note from the above data is that 世界 *shi4jie4,* no matter at which position in a nominal group within a Subject (excluding appearing as a noun head), also served as a modifier within the smaller group. This feature will be examined more later.

7.5 Collocational profile of 世界 *shi4jie4* when occurring within Object

世界 *shi4jie4* had a strong preference for occurring within the grammatical function of an Object (sig.=0.074) compared with the other four comparator nouns. In the sample corpus, there were 268 (out of 535) instances with 世界 *shi4jie4* occurring within the an Object function.

Generally, the verbs preceding 世界 *shi4jie4* when it occurred within an Object were grouped into four semantic associations—ENTER (走向 *zou3xiang4* (to walk into), 走进 *zou3jin4* (to enter) and 进入 *jin4ru4* (to get into)); CHANGE (创建 *chuang4jian4* (to establish), 改变 *gai3bian4* (to change), 建成 *jian4cheng2* (to complete the establishment)); RANK (居 *ju1* (to stand), 跻身 *ji1shen1* (to make one's way to), 位居 *wei4ju1* (to rank), and 占 (to account for)); and FAME (享誉 *xiang3yu4* (to enjoy the fame), 闻名 *wen2ming2* (to be well known)). The following sections examine features of 世界 *shi4jie4* appearing at different positions while severing as part of an Object.

7.5.1 Collocational and semantic associational features when occurring as noun head within Object

The verbs appearing to the left side of 世界 *shi4jie4* appearing as a noun head within an Object could be grouped into a semantic association ENTER, represented by 走向 *zou3xiang4* (to walk to) and 进入 *jin4ru4* (to enter). Unlike *world*, which had a positive colligation with definiteness when appearing as noun head within an Object (with a definiteness to indefiniteness ratio of 4:1), no instance containing 世界 *shi4jie4* belonging to this semantic association occurred with definiteness or indefiniteness.

7.5.2 Collocational and semantic associational features when occurring in premodification

Two patterns of 世界 *shi4jie4* occurred dominantly at the position of a premodification. Similar to 世界 *shi4jie4* occurring within a Subject, 世界 *shi4jie4* occurring within an Object had a positive colligation for occurring with other words to form a combination to premodify the noun head of a nominal group. 世界 *shi4jie4* was likely to participate as part of a premodification in the form of nestings, such as 世界第一 *shi4jie4di4yi1* (the world's best) and 世界各国 *shi4jie4ge4guo2* (every country in the world). Of the 200 instances where 世界 *shi4jie4* appeared as part of a premodification within an Object, 62% (124 out of 200) instances collocated with words associating with a meaning of HIGH RANKING; for example, 世界顶尖

ding3jian1，世界(上)最...*shi4jie4zui4* (the world's best/most/etc.), while 13% (26 out of 200) collocated with words associated with a meaning of WHOLE RANGE OF THE WORLD, represented by the words 各地 *ge4di4* (every corner [of the world]) and 各国 *ge4guo2* (every country [of the world]). Only 3% of instances (six out of 200) occurred in the form of possession 世界的 *shi4jie4d4* (world's).

7.5.3 Semantic association of 认识世界 *ren4shi2shi4jie4* (to understand the world)

It was also interesting to note that 世界 *shi4jie4* also frequently collocated with verbs to form a semi-fixed combination to participate in a premodification*;* for example, 震惊世界 *zhen4jing1shi4jie4* (to surprise the world) *and* 认识世界 *ren4shi2shi4jie4* (to know the world)*.* In combinations of this kind, 世界 *shi4jie4* also served as an Object within the local nominal group. In Chinese grammar, this is called 动宾短语 *dong4bin1duan3yu3*, meaning a phrase made up by Predicate + Object. This phrase can then serve as part of a premodification, as in 震惊世界的举动 (actions which could surprise the world).

The following sections shift focus to the two combinations mentioned above: 认识世界 *ren4shi2shi4jie4* (to understand the world) and 震惊世界 *zhen4jing1shi4jie4* (to shock the world) to examine their distinctive collocational, colligational and semantic associational features. Chapter 5 showed that *understand* (a recurrent translational version of 认识 *ren4shi2*) had a relatively strong collocation strength with *world,* particularly when it served as part of an Object, so data derived from the

Chinese equivalent 认识世界 *ren4shi2shi4jie4* (to understand the world) can be used to detect whether similar combinations with a high MC value in English and Chinese were used in similar ways.

7.5.3.1 Complex issues concerning 认识世界 *ren4shi2shi4jie4* (to understand the world)

There were 3, 137 instances retrieved from the zhTenTen11 corpus, of which 1,000 were extracted for detailed study to a mini-corpus numbered W6. These 1,000 instances included 认识世界 *ren4shi2shi4jie4* (to understand the world) as a fixed combination occurring as part of a premodification. Instances like 我认识世界各国 领导 (I know leaders from countries around the world) were not examined because, in this instance, 世界 *shi4jie4* formed a combination with 各国 *ge4guo2* (every country) rather than being preceded by 认识 *ren4shi* (to understand/know).

Forty-five percent of instances of 认识世界 *ren4shi2shi4jie4* (to understand the world) occurred within an Object as part of a premodification. Of them, 60% (270 out of 450) collocated with another fixed combination 改造世界 *gai3zao4shi4jie4* (to change the world). When we extended our examination to 15 tokens on both sides of 认识世界 *ren4shi2shi4jie4* (to understand the world), we discovered that the combination 认识世界 *ren4shi2shi4jie4* (to understand the world) + 改造世界 *gai3zao4shi4jie4* (to change the world) strongly collocated with words of 马克思主 义 (Marxism) (77 out of 270), 邓小平理论 (Deng Xiaoping Theory) (39 out of

270), and 中国特色社会主义 (socialist culture with Chinese characteristics) (28 out of 270). These words together generalised a semantic association of 认识世界 *ren4shi2shi4jie4* (to understand the world) with CHINESE MARXISM.

The occurrences of 认识世界 *ren4shi2shi4jie4* (to understand the world) +改造世界 *gai3zao4shi4jie4* (to change the world) were more frequent if the examination is not restricted to the Object function; for example:

1) 科学　　是　　**认识世界**　　和　　**改造世界**的

　　ke1xue2　shi4　ren4shi2shi4jie4　he2　gai3zao4shi4jie4de

　　Science　was　**understand the world**　and　**change the world**

　　强大武器。

　　qiang2da4wu3qi4

　　powerful weapon

Science was an effective tool to know the world and a powerful tool to change the world.

While the nesting of 认识世界 *ren4shi2shi4jie4* (to understand the world) had a positive colligation when participating in a premodification within an Object, the combination its English equivalent *understand the world* did not occurred as frequently as 认识世界 *ren4shi2shi4jie4* did in Chinese; only 44 instances were found in the BNC. Sixty-six percent of these instances (29 out of 44) contained *world* occurring within the an Object of *understand*, 17% (seven out of 44) containing *understand the world* occurred within a Subject (e.g. *One way of understanding the world was to see it in terms of the relationships within and between a set of*

interrelated systems…), 9% of instances (four out of 44) containing this combination occurred within a Complement (e.g. *Our purpose was not only to **understand the world**, but to change it*), and the remaining instances occurred in the structure of It was + adjective to do (e.g. *…it was possible to **understand the world** in reductionist terms…*).

It was noted from the comparison between 认识世界 *ren4shi2shi4jie4* (to understand the world) and *to understand the world* that these two combinations were used differently in terms of both collocation and colligation. Chinese speakers' frequent use of 认识世界 *ren4shi2shi4jie4* (to understand the world) was strongly influenced by the unique political situation in China, thus generalising a characteristic usage of this combination.

7.5.3.2 Complex issues concerning 震惊世界 *zhen4jing1* (to shock the world)

This section attempts to classify the 1,296 instances of 震惊世界 *zhen4jing1* (to shock the world) according to their semantic similarities. Three major semantic associations were identified. The general categories of 震惊世界 *zhen4jing1* (to shock the world) and their corresponding sub-classes are revealed in the figure below:

Figure 7.1 Semantic associational distribution of 震惊世界 *zhen4jing1* (to shock the world)

震惊世界 zhen4jing1shi4jie4							
CHANGE					DISCOVERY		DISASTER
NEW POLICY	REVOLUTION	CHANGE	DEVELOPMENT	RIOT	ARCHEOLOGICAL DISCOVERY	SCIENTIFIC DISCOVERY	NATURAL DISASTER
							DISASTER CAUSED BY HUMANS

The first and largest of the semantic associations of 震惊世界 *zhen4jing1* (to shock the world) was CHANGE. Given the size of the group, with a little ingenuity, this semantic association was divided into five sub-classes. The first sub-class referred to NEW POLICY (e.g. 战略防御提议 *zhan4lue4fang2yu4ti2yi4* 'Strategic Defence Initiative'), as illustrated in instance 19 below:

19) 美国总统 里根 宣布 实施

mei3guo2zong3tong3 li3gen3 xuan1bu4 shi2shi1

American president Reagan declare implement

震惊世界的 《**战略防御提议**》。

zhen4jing1shi4jie4de zhan4lue4fang2yu4ti2yi4

shock the world **Strategic Defence Initiative**

The American president Reagan declared that they would implement the SDI, which shocked the world.

The second sub-class referred to a REVOLUTION that contributed to a radical change in a country (e.g. 辛亥革命 *xin1hai4ge2ming4* (The Revolution of 1911) and 科技变革 *ke1ji4bian4ge* (the revolution of science)); instance 20 exemplifies this:

20) 就在这一年， 中国 爆发了

jiu4zai4zhe4yi4nian2 zhong1guo2 bao4fa1le1

In that year	China	broke out

震惊世界的　　　　　　　辛亥革命。

zhen4jing1shi4jie4de　　　xin1hai4ge2ming4

shock the world　　　The Revolution of 1911

In that year, The Revolution of 1911 broke out, which shocked the world.

The third sub-class referred to a TRANSFORMATION that could shock the world (e.g. *改变 gai3bian4/ 变化 bian4hua4* (change)), as illustrated in instance 21:

21) 在过去的　　　二十年里，　　中国的**变化**

zai4guo4qu4de　er4shi2nian2li3　zhong1guo2debian4hua4

In the last　　two decades,　China's change

足以　　　　　**震惊世界**。

zu2yi3　　　　zhen4jing1shi4jie4

enough　　　**shock the world**

China's development in the past twenty years could fairly be able to surprise　the world.

The fourth sub-class concerned DEVELOPMENT (e.g. *贡献 gong4xian4* (contribution) and *成就 cheng2jiu4* (achievement)), as represented in instance 22:

22) 他　　　　　　在数学微积分　　　　领域

ta1　　　　　　zai3shu4xue2wei1ji1fen1　　ling3yu4

He　　　　　　calculus　　　　　　filed

做出了　　　　**震惊世界的**　　　　　**贡献**。

zuo4chu1le1　zhen4jing1shi4jie4de　　　gong4xin4

240

	make	**shock the world**		contribution

In the area of calculus, he had made great contributions that surprise the world.

The fifth sub-class suggested by the data was RIOT (e.g. *政治动乱 zheng4zhi4dong4luan4* (political riots) and *恐怖袭击 kong3bu4xi2ji1* (terrorist attack)); instance 23 shows this usage:

23) 十年前	发生的	那场	**震惊世界的**
shi2nian2qian2	fa1sheng1de	na4chang4	zhen4jing1shi4jie4de
Ten years ago	happen	that	shoced
政治动乱	肇始于	社会	热点问题…
zheng4zhi4dong4luan4	zhao4shi3yu2	she4hui4	re4dian3wen4ti2
political riots	were caused by	social	heated problem

The political riots happened ten years ago was caused by a social heated problem at that time.

The second semantic association of *震惊世界 zhen4jing1* (to shock the world) was DISCOVERY, which was divided into two sub-classes, though the distinction between the sub-classes was not clear cut. The first sub-class referred to an archaeological discovery that may surprise researchers in this field (*兵马俑 bing1ma3yong3* (terracotta warriors)), while the second sub-class referred to scientific discovery, as shown in instance 24:

24) 秦始皇陵	出土的	**兵马俑**
qin2shi3huang2ling2	chu1tu3de	bing1ma3yong3
the tome of Qin Shihuang	unearthed	**terracotta warrior**
震惊世界，		被称为

zheng4jing1shi4jie4	bei4cheng1wei2
shock the world	was credited

世界八大奇迹之一。

shi4jie4ba1da4qi2ji4zhi1yi1

one of the eight miracles in the world

The discovery of terracotta warriors of the First Emperor of China (from 259B.C. to 210B.C.) was awarded as one of the eight miracles in the world.

The third semantic association suggested by the data was that 震惊世界 *zhen4jing1* (to shock the world) associated with semantics expressing DISASTER. This category accounted for 19% of the concordance lines examined. The semantic association with DISASTER could be divided into two sub-classes: natural disaster (e.g. 汶川地震 *wen4chuan1di4zhen4* (Wenchuan Earthquake), 雪灾 *xue3zai1* (snow disaster)) and disaster caused by human beings (e.g. 核辐射 *he2fu2she4* (nuclear radiation), *9.11 事件9.11shi4jian4* (9.11 Incident)). Examples 25 and 26 illustrate these two cases:

25) 去年	5 月 12 日，	我国	发生了
qu4nian2	wu3yue4shi2er4ri4	wo3guo2	fa1sheng1le
Last year	May 2nd	our country	occurred

震惊世界的	汶川	特大地震。
zhen4jing1shi4jie4de	wen4chuan1	te4da4di4zhen4
shock the world	**Wenchuan**	**terrible earthquake**

In May 2nd last year, a terrible earthquake occurred in Wenchuan, China, which shocked the world.

26) 25 年后，	日本福岛的	核泄漏
er4shi2wu3nian2hou4	Fukushima Island	nuclear leakage

事故	又一次	震惊世界。
accident	another time	shock the world

25 years later, the effects of nuclear leakage accident which had occurred on Fukushima island shocked the world again.

Examination of 震惊世界 *zhen4jing1* (to shock the world) showed that the notion of nesting as defined in Lexical Priming Theory was also valid in Chinese. As a product of priming, the nesting itself was primed in a distinctive way, in that it was associated with different meanings and colligated positively within a particular grammatical function.

Next, we investigated the semantic associations of 震惊世界 *zhen4jing1* (to shock the world) by looking at its more distant neighbours. Another combination 世界震惊 *shi4jie4zhen4jing1* (world shocks) was examined evidence semantic differences in cases of changed grammatical position.

The semantic association with CHANGE accounted for 44% (570 out of 1296) of the total instances of 震惊世界 *zhen4jing1shi4jie4* (to shock the world). Of these, 11% (63 out of 1296) constituted the first sub-class (NEW POLICY). Extending our examination to 15 tokens to both sides, we noted that the textual association with this sub-class expressed a relationship of statement — criticism. The new policy mentioned in the context was likely to shock the world in a negative way, or worse, bring pain to people in the world; instance 19 exemplifies a new policy (Strategic Defence Initiative).

Eighteen percent (103 out of 1296) of instances formed the second sub-class of the

CHANGE semantic association — REVOLUTION. Ninety-six percent of these cases (99 out of 103) expressed the significance of revolution; for example, instance 20 referred to one of the most influential revolutions in Chinese history — the 1911 Revolution — which to a large extent marked the beginning of modern China as feudalism was overthrown by the revolutionists led by Sun Yat-sen.

Regarding the third sub-class of CHANGE — CONTRIBUTING TO THE DEVELOPMENT OF THE WORLD — almost 100% of the cases displayed a textual association with emphasizing China's influence on the world and China's efforts to promote the world's development. The boundary between this sub-class and the DISCOVERY category was rather blurred and not watertight, as discoveries that could shock the world, including archaeological and scientific ones, could also update researchers' knowledge in a specific field and surprise them. This pragmatic association was evidenced by the positive attitude expressed in concordance lines belonging to the semantic association with DISCOVERY.

With respect to the fifth sub-class of CHANGE, we have found the dominant textual function in the context was one of boosting the credit afforded the government. This textual association was also evidenced by one of the sub-classes setting up the semantic association with DISASTER — NATURAL DISASTER — which accounted for 78% of total occurrences (303 out of 389). The dominant members of the semantic set were 汶川地震 *wen4chuan1di4zhen4* (Wenchuan earthquake) and 雪灾 *xue4zai1* (snow disaster). The pragmatic association indicated in the context tended to praise the government's quick reaction to disasters and excellent capability for implementing plans for saving survivors, comforting victims, and helping locals

re-build their homes.

However, when talking about disasters caused by human beings, the attitude of the context tended to be complaining and criticising. Eighty-six instances of this sub-class were examined. Interestingly, 84 of these "surprising disasters caused by human beings" occurred in foreign countries, such as the 福岛核辐射 *fu2dao3he2fu2she4* (Fukushima Radiations) that occurred in Japan and the 踩踏事件 *cai3ta4shi4jian4* (stampede) that occurred in Cambodia in 2010; the two exceptions described the Nanjing Massacre committed by Japanese troops and an explosion at a chemical factory.

The above data shows that no, matter what semantic association 震惊世界 *zhen4jing1shi4jie4* (to shock the world) was occurring with, it was likely to express three categories of textual semantic associations—emphasising China's contribution to the world's development, praising the government, and criticising those who have done harm to ordinary people. The majority of instances contributed to the first two categories, implying native Chinese speakers are typically primed to use the combination 震惊世界 *zhen4jing1shi4jie4* (to shock the world) to show their pride in China's contributions to the world and their support for the government.

Another representative structure in premodifications within an Object was Subject (shi4jie4) + Predicate. The typical right verbs 震惊 *zhen4jing1* (to surprise), occurring 233 times, and 惊叹 *jing1tan4* (to be astonished or to be amused), occurring 149 times, formed a semantic set associated with a meaning of

SURPRISE. The textual semantic association suggested by the combination 世界震惊 *shi4jie4zhen4jing1* (to shock the world) was to emphasise China's rapid development and fantastic performance at the Beijing Olympic Games, whereas the combination 世界惊叹 *shi4jie4jing1tan* (to make the world astonished or to amuse the world) typically had a discourse function of crediting Chinese people's intelligence, crediting the government's efforts to make "miracles" (for example, helping rescue earthquake survivals within a very short time) and crediting the significant developments Chinese people have achieved. The data suggest Chinese speakers are likely to express a positive attitude towards the Chinese people and government through the combination 世界震惊 *shi4jie4zhen4jing1* (to shock the world).

Preliminary evidence suggests Chinese is likely to be primed for textual semantic association, both in Hoey's original sense and in our extended sense. As we have seen from our Chinese data, textual semantic association can be manifested through nestings. In cases where component order changed in a nesting, the textual priming also changed accordingly.

7.6 Collocational profile of 世界 *shi4jie4* when occurring within Complement

世界 *shi4jie4* also avoided appearing within a Complement, with only 37 of its 535 instances occurring within this function. To obtain sufficient data, we retrieved 500 instances from zhTenTen11 in which 世界 *shi4jie4* occurred within a Complement. Of these instances, 78% (390 out of 500) contained 世界 *shi4jie4* as part of a

premodification while the remaining 22% instances contained 世界 *shi4jie4* as a noun head. The most frequent combination containing 世 界 *shi4jie4* in a premodification was 世界上 *shi4jie4shang4* (in the world) (54%). The second most frequent combination (28%) was realised in the structure of 世界 *shi4jie4* + Noun + 最. As noted in Chapter 6, the Chinese character 最 *zui4* is equivalent to many Superlatives in English when followed by adjectives, such as 最大 *zui4da4* (the biggest/largest), 最好 *zui4hao3* (the best), and so forth. Apart from these two kinds of collocates, 世界 *shi4jie4* also co-occurred with words expressing polar meanings, such as 著名 *zhu4ming2* (famous), 顶尖 *ding3jian1* (top) and 第一 *di4yi1* (first), and co-occurred with words expressing PLACE, such as 各国 *ge4guo2* (every country) and 各地 *ge4di4* (everywhere) to form a local modifier in a premodification within a Complement.

As when occurring within an Object, 世界 *shi4jie4* also had a positive preference for participating into a premodification in the form of a word sequence (or nesting) when occurring within a Complement, where 世界 *shi4jie4* could appear as an Object (e.g. 震惊世界 *zhen4jing1shi4jie4* (to surprise the world)), as a Subject (e.g. 世界震惊 *shi4jie4zhen4jing1* (to make the world astonished)), or as a Modifier (e.g. 世界各国 *shi4jie4ge4guo2* (every country in the world)). Compared with English, Chinese has more various forms in a nominal group, including classifier phrases, measure phrases, associative phrases, modifying phrases, verb phrase, etc. With respect to categories of phrases, Chinese can be divided into nominal phrases, verbal phrases, subject-predict phrases, verb-ubject phrases, etc. The complicated phrases may

confuse learners of Chinese, especially those from English-speaking countries. In addition, its complicated word structure also makes it difficult for cross-linguistic researchers to identify units for comparison. In terms of defining equal units for contrast linguistic studies, the data from this study suggest that nesting could be applied as a corresponding unit of meaning across languages. Cross-linguistic equivalence might be realised through a nesting with similar semantic and textual primings.

7.7 Collocational profile of 世界 *shi4jie4* when occurring within Adjunct

Chapter 6 showed that 世界 *shi4jie4* did not have a clear preference or aversion for occurring within the grammatical function of Adjunct, compared to other nouns; yet, it had a strong preference for occurring in two combinations when functioning as an Adjunct. Of the 107 instances containing 世界 *shi4jie4* within an Adjunct, 58 appeared as part of a premodification, while 49 instances occurred as a noun head.

Within the grammatical position of a premodification, 33% of instances occurred in the form of 世界上(中 *zhong1*, 里 *li3) shi4jie4shang4* (in the world), 51% occurred in the combination *(在)*世界范围内 *zai4shi4jie4fan4wei2nei4* (in the range of the world), 9% in the combination 第一(二)次世界大战 *di4er4ci4shi4jie4da4zhan4* (the first/second world war)*, and 7% in the combination of 从 *cong2* (from…) with 世界 *shi4jie4+ noun head* 来看 *lai2kan4* (to see); for example, 从世界经验来看 (viewed from the perspective of world experience…)

Of the 49 instances containing 世界 *shi4jie4* appearing as a noun head in the

combination participating in a premodification, a dominant structure was with 在 *zai4* + POSSESSIVE PRONOUN (e.g. 我的 *wo3de my*, 你的 *ni3de your*) + 世界 *shi4jie4*, thus indicating a colligational preference for co-occurring with possessive pronouns within an Adjunct.

7.8 Collocational behaviour of nestings containing 世界 *shi4jie4*

The structure of pronoun plus the morpheme 的 *-de* forms a typical type of modification belonging to associate phrases in Chinese. As the name suggests, in this kind of phrase two noun phrases are "associated" or "linked" in some way. The meaning of the connection is determined by the meanings of the two noun phrases involved. Possession is an important associative meaning, such as 我的衣服 *wo3deyi1fu2* (my clothes). However, there are also many other possible associations between two noun phrases where the semantic nature of the association follows from the meanings of the two noun phrases involved, such as the example (revealed in our data) 我的世界 *wo3deshi4jie4* (my world). Sometimes, the *-de* in an associative phrase signalling a possessive relationship can be omitted when the relationship is between two human relatives; for example, 我（的）妹妹 *wo3(de)mei4mei* (my younger sister). It is not necessary to distinguish every possibility of the connection in associative phrases in Chinese and it seems very demanding for learners of Chinese (particularly learners from English-speaking countries) to apply the correct structure fitting each possibility. From a pedagogical perspective, we may need to account for Chinese phrase structure from the perspective of learners of Chinese. The conceptions of nesting and priming could be referenced to account for the complex

phrase structures of Mandarin, assuming phrase structure is realised through a combination nesting. As a priming product of the individual word constituting it, a nesting is supposed to be primed with a particular word or nesting within a particular colligational pattern. This assumption is valid only if we have proved the features of nesting, as defined in Lexical Priming Theory, can be found in Chinese. Thus, this section will examine the behaviour of nestings in Chinese.

This section investigates the collocational features of frequently used nestings containing 世界 *shi4jie4* to test whether the semantic associations of a combination and its components differ from each other in Chinese. By doing so, we intended to find evidence to support our modification of Hoey's definition of nesting — i.e., that the collocational and semantic associational behaviour of a nesting and its components have similarities; however, they may differ in terms of collocational strength and semantic set members.

The five most frequently used nestings drawn from the zhTenTen11 corpus are listed below, with their times of occurrences provided in the bracket.

世界经济 *shi4jie4jing1ji4* (the world economy) *(30,978)*

成为世界上 *cheng2wei2shi4jie4shang4* (become … in the world) *(17,355)*

在世界范围内 *zai4shi4jie4fan4wei2nei* (in range of the world) *(13,591)*

走向世界 *zou3xiang4shi4jie4* (walk to the world) *(13,380)*

第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) *(11,750)*

Our research procedure was as follows:

- Examine the semantic associations of these five nestings and compare them with those of the individual word 世界 *shi4jie4*.

- Examine the semantic associations of their English equivalent nestings (determined by MC value) and compare the semantic associational behaviours of the corresponding Chinese and English nestings.

- Examine the semantic associational features of Chinese equivalents to the most frequently used English nestings containing *world* and compare the behaviours found for Chinese and English.

7.8.1 Semantic association of 世界经济 *shi4jie4jing1ji4* (the world economy)

Twenty-six instances of 世界经济 *shi4jie4jing1ji4* (the world economy) were retrieved from the LCMC and placed in a database named DW11. The instances indicated three major semantic associations of the combination understudy. The first was BEING DEPENDENT, with representative collocates of 集团化 *ji2tuan2hua4* (collectivisation) and 依赖 *yi1lai4* (depend on) (exemplified in instance 26); the second was DEVELOPMENT, constituted by semantic set including 增长 *zeng1zhang3* (increase), 发展 *fa1zhan3* (develop), and 增长率 *zeng1zhang3lv* (growth rate) (indicated in instance 29); and the third was UNIT OF TIME, represented by the semantic set members *80 年代 ba1shi2nian2dai4* (1980s) and *90 年代 jiu3shi2nian2dai4* (1990s) (as exemplified in instance 29):

27) 世界经济　　　　**集团化**　　　趋势　　　可能

shi4jie4jing1ji4	ji2tuan2hua4	qu1shi4	ke3neng2
world economy	**collectivization**	trend	may
导致	贸易保护主义		加强。
dao3zhi4	mao4yi4bao3hu4zhu3yi4		jia1qiang2
cause	trade protectionism		strengthen

The trend of world economy collectivization may strengthen the trade protectionism.

28) 世界经济　　　　　　正在　　　　　　蓬勃发展。

shi4jie4jing1ji4　　　zheng4zai4　　　peng2bo2fa1zhan3

The world economy　　　was　　　developing vigorously.

The world economy was developing vigorously.

29) **90 年代，**　　　　世界经济的　　　　　增长率

90nian2dai4　　　shi4jie4jing1ji4de　　　zeng1zhang4lv4

1990s　　　　world economy　　　increasing rate

将会　　　　　低于　　　　　80 年代。

jiang1hui4　　　di1yu2　　　80nian2dai4

will　　　lower　　　1980s

The increasing rate of the world economy in 1990s will be lower than that of 1980s.

The above is a general semantic profile of the nesting 世界经济 *shi4jie4jing1ji4* (the world economy). We discovered in the previous chapter that a combination may share some collocates and semantic associations with its components; thus, we intended to modify Hoey's definition of nesting by claiming that a nesting and its components do have shared collocates and semantic associations, but differ in the

collocational strength of co-occurring with a word and semantic members attributing to the shared semantic association; we assume this modification is also valid for Chinese. We researched the word 经济 *jing1ji4* (economy) to detect whether this individual word shared collocates or semantic sets with the nesting containing it.

Sketch Engine showed 2,694,296 counts of 经济 *jing1ji4* (economy), with 1,278.90 occurrences per million running words. The collocates at the L1 position of 经济 *jing1ji4* (economy) are displayed table 7.1 in descending order of LogDice score:

Table 7.1 L1 collocates of 经济 *jing1ji4* (economy)

collocates		Frequency	LogDice score
市场	market	110,653	9.065
国民	national	60,032	9.444
循环	sustainable	36,974	8.745
区域	region	33,236	8.451
宏观	macro	27,414	8.318
中国	China	58,205	8.174
世界	world	30,979	8.071
农村	rural	29,246	8.048
地方	local	27,475	8.038
社会	society	46,875	8.036
我国	our country	29,225	8.032
民营	enterprise	20,852	7.93
工业	industry	23,993	7.889
促进	promote	23,560	7.788
全球	global	20,332	7.778

This table, though not able to indicate the L1 collocates of 经济 *jing1ji4* (economy) in any depth, shows the most highly scored L1 collocates. The word 世界 *shi4jie4* ranked seventh in the table. Examination of these collocates revealed that semantic associations of 经济 *jing1ji4* (economy) can be grouped into three different

categories. The first and largest was associated with REGION/AREA (区域 *qu1yu4* (region), 中国 *zhongguo2* (China), 世界 *shi4jie4*, 农村 *nong2cun1* (rural area), 地方 *di4fang1* (local), 我国 *wo3guo2* (our country), 全球 *quan2qiu2* (global)). The other two were ECONOMY TYPES (国民 *guo2min2* (national), 循环 *xun2huan2* (sustainable), 宏观 *hong2guan1* (macro), 社会 *she4hui4* (society), 民营 *min2ying2* (enterprise), 工业 *gong1ye4* (industry)) and DEVELOPMENT (促进 *cu4jin4* (to promote), 提高 *ti2gao1* (to increase), 发展 *fa1zhan3* (to develop)). There were several collocates shared by 世界 *shi4jie4* and 经济 *jing1ji4* (economy), such as 市场 *shi4chang3* (market) and 当前 *dang1qian2* (today), which appeared on a later page with a LogDice score of 5.964, a lower strength than with 世界 *shi4jie4*, which had a LogDice score of 7.79. There were also collocates shared by 经济 *jing1ji4* (economy) and the combination 世界经济 *shi4jie4jing1ji4* (world economy), including 提高 *ti2gao1* (to increase) and 发展 *fa1zhan3* (to develop), although those were not shared with the component 世界 *shi4jie* (world). This combination, however, shared other collocates with 世界 *shi4jie* (world), such as 当今 *dang1jin1* (today).

The examination of the combination 世界经济 *shi4jie4jing1ji4* (world economy) and its two components also showed that, as was shown for *world* in Chapter 4, the combination and its component may share semantic associations. The shared semantic association of 世界经济 *shi4jie4jing1ji4* (world economy) and 世界 *shi4jie* (world) was UNIT of TIME, and the shared semantic association of 世界经济 *shi4jie4jing1ji4* (world economy) and 经济 *jing1ji4* (economy) was DEVELOPMENT. The difference was that the shared semantic association was not

represented by the same semantic set members. Another result derived from this examination was that, as with its English equivalent, 世界 *shi4jie* (world) shared collocates with the combination containing it; however, these shared collocates may differ in their strength of co-occurrence.

7.8.2 The semantic associations of 成为世界上 *cheng2wei2shi4jie4shang4* (become…in the world)

The second combination examined was 成为世界上 *cheng2wei2shi4jie4shang4* (become…in the world), which can be separated into two components: 成为 *cheng2wei2* (become) and 世界上 *shi4jie4shang4* (in the world). This phrase could also be translated into different versions, such as *around the world* and *world's*; here, we use the version with the highest MC[20] value *in the world* (36.2%[21]), but that does not mean it was a universal equivalent to 世界上 *shi4jie4shang1* in all contexts.

We began bey examining the phrase 世界上 *shi4jie4shang1*. The English equivalent to this combination or phrase was used rather frequently (7,236 occurrences) in the BNC, so comparing the two combinations (世界上 *shi4jie4shang1* vs. *in the world*)

---

[20] MC value: Mutual Correspondence value, which is used as a measure to translatability between source language and its target translation, which helps to give an indication of the degrees of correspondence between two words (source and target). This value has been specified in detail in Section 3.2.3, please refer to this chapter for detailed information.

  MC value: Mutual Correspondence value, which is used as a measure to translatability between source language and its target translation, which helps to give an indication of the degrees of correspondence between two words (source and target). This value has been specified in detail in Section 3.2.3, please refer to this chapter for detailed information.

  in size, so we reduce the reference High MC value to 30% and Low MC vale to 5%. . Those with an MC of 30% or higher is regarded as a High MC Group, and a Low MC Group if the value is 10% or below.

provided distinctive data. It was presumable that they may have a number of differences in terms of collocation and semantic association because, as discussed in Chapter 3, English and Chinese are distinctively different languages, making their similarities more interesting and worth exploring. If 世界上 *shi4jie4shang1* and *in the world* had similarities, it would show that both English speakers and Chinese speakers shared an understanding and interpretation of the combination 世界上 *shi4jie4shang1* and *in the world*. We then further hypothesised that priming can account for why speakers from both countries unintentionally reproduce this combination in this situation, sometimes in the same way.

The combination 世界上 *shi4jie4shang1* was divided into two categories. One included instances that could not be treated as equivalent to *in the world*:

30) 我们　　来到　　这个　　**世界上**

wo3men2　lai2dao4　zhe4ge4　shi4jie4shang4

We　　come to　this　　**world**

不是　　为了　　　　　　捡面包屑

bu2shi4　wei4le1　　　　jian3mian4bao1xie4

not　　for　　　　　　pick up crumb

和　　　烂骨头，　　我们　　向往

he1　　lan4gu3tou2,　wo3men2　xiang4wang3

and　　rotten bones　we　　yearn

到大海去。

dao4da4hai3qu4

go to the see

We came to the world for running toward the sea, rather than picking up crumbs and rotten bones.

where 世界上 *shi4jie4shang1* had a similar meaning with the individual word 世界 *shi4jie4* and the particle 上 *shang4* had no particular meaning. *Shang4* in Chinese is a word of directional locality, such as, *shang4mian4* (upside or on top of something). When used as an adjective, it refers to something mentioned above or previously; for example, 上文 *shang4wen2* (the abovementioned chapters). In numerous cases, it is used after an action verb to express direction, such as 走上去 *zou4shang4qu4* (to go up) or 跑上去 *pao3shang4qu4* (to run up). However, when it collocated with 世界 *shi4jie4* and placed itself on the right side, it lost its independent meaning and had nothing to do with direction. It formed a unity with 世界 *shi4jie4* as a whole, being largely equivalent to *in the world* in English. In case 38, the *shang4* in the combination 世界上 *shi4jie4shang4* could be omitted without loss of meaning; thus, the meaning of 世界上 *shi4jie4shang4* in this case matches that of the English nesting, *in the world*. In this sense, this combination can be classified into two categories—instances where 世界上 *shi4jie4shang4* had the same meaning of *in the world*, and instances where it did not.

There were 102,151 instances of 世界上 *shi4jie4shang4* found in the zhTenTen11 corpus. As was done previously, we built a database containing 1000 instances by retrieving five instances from each of the first 200 concordance pages from different websites[22].

---

[22]zheTenTen11 is a corpus made up with online articles, so the sources of each concordance line

There were 20 instances where 世界上 *shi4jie4shang4* did not have a meaning equivalent to *in the world*. The combination 世界上 *shi4jie4shang4* in 20 examples could be replaced by the individual word 世界 *shi4jie4*, where *shang4* did not affect the sense of the combination. Seventy-five percent (15 out of 20) of these instances contained 世界上 *shi4jie4shang4* in an Object (exemplified in instance 38) with the remaining instances of 世界上 *shi4jie4shang4* functioning as a Subject, illustrated in example 31 below:

31) 中国　　　　　乃至　　　　　**世界上**

　　　zhong1guo2　　nai3zhi4　　　shi4jie4shang4

　　　China　　　　even　　　　　**the world**

　　　都是　　　　　绝无仅有的。

　　　dou1shi4　　　jue2wu2jin3you3de

　　　were all　　　one-off

[It] was a one-off in China and even in the world.

The remaining 980 instances containing 世界上 *shi4jie4shang4* had an equivalent meaning to *in the world*. The collocates in the L1 position of this combination included 是 *shi4* BE (392 hits, 40% of the extracted instances), 当今 *dang1jin1* (today) (98 hits, 10%), 目前 *mu4qian2* (at present) (49 hits, 5%), 当前 *dang1qian2* (now) (29 hits, about 3%), 走访 *zou3fang3* (to visit and have an interview) (69 hits, about 7%), 作为 *zuo4wei2* (to be treated as) or functioning as a preposition equivalent to *as* (48 hits, about 5%), and 成为 *cheng2wei2* (to become) (46 hits,

---

is named after the source web where it belongs to.

about 5%). These collocates then formed three major semantic associations: TIME (*当今 dang1jin1* (today), *目前 mu4qian2* (at present), *当前 dang1qian2* (now)); VISITING (*走访 zou3fang4* (to meet and have an interview); *拜访 bai4fang3* (to visit)) and BECOME (*成为 cheng2wei2* (to become), *变为 bian4wei2* (change into)). However, these three dominant associations do not include the most frequent L1 collocate, i.e., *是 shi4* BE, which accounted for the largest proportion of all the collocates occurring at this position. The collocation of *是 shi4* BE + *世界上 shi4jie4shang4* were more likely to have a semantic association with EVALUATION, which tended to be followed by words such as *最 zui4* (the most or the best), *第一 个 di4yi1ge4* (the first one), or *唯一 wei2yi1* (the only one):

32) 荷兰　　　是　　　世界上　　　　　**最大的**

　　he2lan2　shi4　shi4jie4shang4　zui4da4de

　　Holland　was　in the world　　**biggest**

　　鲜花　　　输出　　　　　　　　国。

　　xian1hua1　shu1chu1　　　　　　guo2

　　flower　　export　　　　　　　country

Holland was the biggest flower exporting country in the world.

This section investigates collocates at the R1 position of *世界上 shi4jie4shang4*. The dominant collocates found were *最 zui4* (the most or the best) (150 hits, 15% of 980 recovered instances), *的[23] de1* ('s) (a possessive marker in Chinese) (109 hits, 11%),

---

[23] *的 de1* in Chinese is a structure word. Sometimes it forms the end of adjective (e.g., *漂亮的 piao4liang4de* (beautiful)), sometimes the end of a sentence (*好的 hao3de1* (yes)), and

第一 *di4yi1* (the first) (98 hits, 10%), 各种 *ge4zhong3* (every kind of) (23 hits, 2.3%), 任何 *ren4he2* (any) (30 hits, 3%), 大多数 *da4duo1shu4* (the majority of) (57 hits, 6%), and 引起 *yin4qi3* (to cause) (63 hits, about 6%). These collocates formed a dominant semantic association — EMPHASIS — which also had a pragmatic association with EVALUATION:

33) 1957 年　　　　　　世界上　　　　　　第一颗

　　　1959nian2　　　　shi4jie4shang4　　di4yi1ke1

　　　In 1957　　　　　**in the world**　　**the first**

　　　人造卫星　　　　　发射　　　　　　　成功。

　　　ren2zao4wei4xing1　fa1she4　　　　cheng2gong1

　　　satellite　　　　　launch　　　　　successfully

In 1957, the first satellite in the world was launched successfully.

Having examined the lexical behaviour of 世界上 *shi4jie4shang4* at the L1 and R1 positions, we then examined the semantic associational behaviour of another component of 成为世界上 *cheng2wei2shi4jie4shang4* (become…in the world) — 成为 *cheng2wei2* (to become). The representative L1 collocates of 成为 *cheng2wei2* (become) were 如今 *ru2jin1* (current), 目前 *mu4qian2* (at present), 将来 *jiang1lai2* (the future), 现在 *xian4zai4* (now), 建设 *jian4she4* (establishment), 产业 *chan3ye4* (industry), 我国 *wo3guo2* (our country), 旅游业 *lv3you2ye4* (tourism), 互联网 *hu4lian2wang3* (the Internet), etc. The most frequent R1 collocates of 成为

---

sometimes, it functions as a marker of possession (e.g., 我的 *wo3de1* (my mine) or 世界的 *shi4jie4de1* (the world's).

cheng2wei1 (become) were 焦点 *jiao1dian3* (focus) (7.96), 之一 *zhi1yi1* (one of the…) (7.86), 世界 *shi4jie4* (7.81), 热点 *re4dian3* (heated issue) (7.12), and 国内 *guo2nei4* (domestic) (6.99); less frequent R1 collocates are 中国 *zhong1guo2* (China) (6.21), 领域 *ling3yu4* (area) (6.11), and 我国 *wo3guo2* (our country) (6.48). The dominant semantic association was constituted by a semantic set expressing AREA, such as 世界 *shi4jie4,* 中国 *zhong1guo2* (China)*, and* 全球 *quan2qiu2* (the whole globe)*,* etc. In addition to these frequently occurring collocates, some collocates with lower LogDice scores also belonged to the semantic set forming AREA, such as 领域 *ling3yue4* (area) (4.301). This finding reinforces the assumption that the concept of semantic association defined in Lexical Priming Theory also applies to Chinese. Collocation accounts for the co-occurrences appearing together in corpora more often than random distribution, while semantic association accounts for the phenomenon that when a word or word sequence is associated with a semantic set, words forming the set may have various collocational strengths (or none) of occurring with the node word.

The above paragraphs examined the collocational and semantic associational features of components making up the combination 成为世界上 *cheng2wei2shi4jie4shang4* (become…in the world). To investigate whether this combination shared semantic associations with its components, a study sketching collocates and categorising the semantic associations of this combination was carried out. There were 17,355 instances in the concordance from the corpus and a database comprising 1000 instances from each kind of website was set up (DW 13). The left-hand semantic association with this combination was COUNTRY, dominantly represented by

CHINA (e.g. *中国 zhong1guo1* (China), *我国 wo3guo2* (our country), *日本 ri4ben3* (Japan)) and ESTABLISHMENT (e.g. *建 成 jian4cheng2* (completing the establishment), *建设 jian4she4* (construction)), whereas the right-hand semantic association was dominantly represented by EVALUATION (e.g. *第三大 di4san1da4* (the third biggest) and *一流 yi1liu2* (first rate)). There were shared semantic associations to the left and right of the word sequence *成 为 世 界 上 cheng2wei2shi4jie4shang4* (become…in the world), although members of the semantic set varied to different degrees.

This finding, echoing that presented in Chapter 4 for *world*, shows it is possible for a word sequence to have shared collocates and semantic associations with its components; however, members of the semantic set of shared semantic associations might differ.

7.8.3 The semantic association of *走向世界 zou3xiang4hsi4jie4* (to walk to the world)

In this section, the fourth nesting, *走向世界 zou3xiang4hsi4jie4* (to walk to the world), is investigated. The third most frequently used nesting *在世界范围内 zai4shi4jie4fan4wei2nei4* (in the range of the world) will be discussed in the next chapter as it tended to have various grammatical functions and its semantic association varied when serving different functions, meaning it would be more appropriate to discuss its semantic associations by taking its collocational behaviour into account.

While the literal translation provided for 走向世界 *zou3xiang4hsi4jie4* is *to walk to the world*, its wider meaning is that something is open to the world and is becoming famous and influential in the world or in a specific area; for example:

3442) 在　　国内品牌　　　还不是　　　很成熟的时候，

zai4　　guo2nei4pin3pai2　　hai2bu4shi4　　hen3cheng2shu2deshi2hou4

When　　domestic brand　　was not　　very mature

健力宝　　　已经　　　　　　　　走向世界。

jian4li4bao3　　yi3jing1　　　　　　zou3xiang4shi4jie4

Jianlibao　　had already　　　　　walked to the world

When domestic brand was still not mature, the company of Jianlibao[24] had already become international.

In the above example, 走向世界 *zou3xiang4shi4jie4* obviously implies that the *Jianlibao* brand is known to the world in the field of energy drink production. We used *to walk to the world* as a general literal translation of 走向世界 *zou3xiang4shi4jie4* for the purpose of simplification.

The collocates to the left and right side were extended to five words because a majority of the left collocates of 走向世界 *zou3xiang4hsi4jie4* (to walk to the world) were "tense markers" (Chinese does not strictly have tense in its grammar, as previously discussed; here, "tense markers" refers to words or morphemes indicating time duration), such as 即将 *ji2jiang1* (will), 已经 *yi3jing1* (have/had), and so forth.

---

[24] An influential beverage brand in China. This company was founded in 1984 and has many sub-brands. Its businesses includes drinks, seasonings, and housing industry.

These words all showed a colligational preference for a particular tense; however, this does not clearly depict a semantic association. They did display a set of words that can be grouped into a particular category, but this category was more likely to grammatical than semantical.

Other instances were extracted to form the DW13 database. Of these 1000 instances, 80% occurred at the end of a clause, which was a clear textual colligation. Of the remaining 200 instances that occurred in other positions of a clause, 95% were followed by the morpheme 的 *de*, a typical particle linking two relative clauses or phrases to form an adjective; for example, 走向世界的桥梁 (a bridge to the world), where 的 *de* links the nesting of 走向世界 *zou3xiang4shi4jie4* (to walk to the world) and the noun head 桥梁 qiao2liang2 (the bridge). In the preceding combination, 走向 *zou3xiang4* (to walk to) was a verb while 世界 *shi4jie4* was its Object. In Chinese grammar, this phrase is called 动宾短语 *dong4bin1duan3yu3* (verb-Object phrase). A verb-Object structure can also be seen as a complete structure in a Chinese clause, so this combination tended to occurred at the end of a clause because it had, to a large extent, accomplished the mission of expressing a complete meaning. The following paragraphs investigate the 800 clause-final instances containing the combination 走向世界 *zou3xiang4shi4jie4* (to walk to the world).

Three semantic associations were categorised. The largest one was CHINESE FEATURES (622 instances, accounting for 78% of all instances examined), which had four sub-classes: CHINA (中国 *zhong1guo2* (China), 我国 *wo3guo2* (our

country)); INSTITUTIONS in China (*高校 gao1xiao4* (universities), *金融机构 jin1rong2ji1gou4* (Financial Institution), *中国企业 zhong1guo2qi3ye4* (Chinese entrepreneurs)), Chinese CULTURE (*中国文化 zhong1guo2wen2hua4* (Chinese culture), *历史遗存 li4shi3yi2cun2* (historical heritage), *中医 zhong1yi1* (Chinese medicine)); and Chinese ARTS (*中国电影 zhong1guo2dian4ying3* (Chinese films), *红楼梦 hong2lou2meng4* (A Dream of Red Mansions)[25]). Of the four sub-classes, Chinese CULTURE accounted for the largest proportion (39%), while CHINA accounted for the smallest (11%).

The second semantic association detected from our data was DEVELOPMENT (represented by *发展 fa1zhan3* (development), *改革 gai3ge2* (reform)). Eighty members of the semantic set formed this semantic association (10% of all instances). The third and smallest semantic association was YOUNGER GENERATION (represented by *年轻一代 nian2qing1yi1dai4* (younger generation) and *新一代设计师 xin1yi1dai4she4ji4shi4* (the designer of a new generation)).

Revisiting the general semantic associational profile of *世界 shi4jie4*, we found *世界 shi4jie4* and the combination *走向世界 zou3xiang4shi4jie4* (to walk to the world) shared a semantic association with COUNTRY. Interestingly, this semantic association occurred with the same percentage (11%) among all instances under study.

7.8.4 Semantic association of *第二次世界大战 di4er4ci4shi4jie4da4zhan4* (the

---

[25] A remarkable novel about Chinese feudal society. It enjoys a high reputation in both China and abroad.

second world war)

The last frequent nesting examined in this section is 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war). In total, there were 11,705 instances occurring in the zhTenTen11 corpus, with 5.60 instances per million running words.

Two major semantic meanings were associated with 大战 *da4zhan4* (big war). These semantic associations were distinguished according to the sense chosen by 大战 *da4zhan4* (big war); i.e., practical war (denotative meaning) or figurative war (connotative meaning). When 大战 *da4zhan4* (big war) was associated with a real war it collocated with 世界 *shi4jie4*, 台儿庄 *tai2er2zhuang1* (the name of a small village in China), 中原 *zhong1yuan2* (a general name for areas covering the middle and lower reaches of the Huanghe River, China), and 赤壁 *chi4bi4* (or Chihpi, situated on the south bank of the Yants River, Hupeh Province, China), forming a semantic association with the PLACE in which the war happened. However, when 大战 *da4zhan4* (big war) was associated with a figurative meaning — i.e., a fierce competition — it was more likely to be preceded by words such as 点球 *dian3qiu2* (penalty kick), 坦克 *tan3ke4* (tank), 人机 *ren2ji1* (human beings and machines), 生源 *sheng1yuan2* (sources of students), and 口水 *kou3shui3* (spittle) (referred to a fight or a competition of words when collocating with 大战 *da4zhan4* (big war)), which then formed a semantic association relating to two sides of a competition.

The semantic association with 大战 *da4zhan4* (big war), however, did not affect the

266

semantic profile associated with 世界大战 *shi4jie4da4zhan4* (the world war). There were only two words appearing at the L1 position to this combination—*first* and *second*—with the frequency of the latter outweighing that of the former. The higher possibility for 世界大战 *shi4jie4da4zhan4* (the world war) occurring after the words *first* and *second* might be determined by the other component of this combination—世界 *shi4jie4*—because ordinal words formed a strong semantic association with NUMBER to the L1 position of 世界 *shi4jie4.*

Since 第二次 *di4er4ci4* (second) had a higher collocational strength (7.997) than that of 第一次 *di4yi1ci4* (first) (6.143) when collocating with 世界大战 *shi4jie4da4zhan4* (world war), therefore, the combination 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) appeared far more frequently than 第一次世界大战 *di4yi1ci4shi4jie4da4zhan4* (the first world war). An examination of the retrieved 1000 instances (DW14 database) containing 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) showed that the combination had a strong preference for a particular position in a sentence or text. Seventy-five percent of the extracted instances in DW14 had 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) occurring at the beginning of a clause, functioning as part of an Adjunct. The dominant collocate appearing in front of this combination was 在 *zai4*, which can be translated to a number of prepositions in English, such as in, at, during, since, and so on, depending on the words occurring after this combination. For example, 在第二次世界大战期间 *zai4di4er4ci4shi4jie4da4zhan4qi1jian1* (**during** the second world war, where 在 *zai4* equals to preposition of *during*) accounted for 44% (328) of total instances in

this condition. Another dominant collocate preceding 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) when it occurred at the beginning of a clause was 从 *cong2* (since or from..., again depending on the words subsequent to this combination), accounting for 15% of the data (113). The remaining instances containing the word sequence were only followed by right collocates when occurring at the initial position in a clause. These collocates then constituted a semantic set of expressing TIME PERIOD, such as 结束 *jie2shu4* (end), 期间 *qi1jian1* (duration), 中 *zhong1* (middle), and so forth. Collocates forming this semantic association accounted for 88% of all remaining instances.

Apart from the 750 instances of 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) appearing at the beginning of a sentence while functioning as an Adjunct, there were 250 instances occurring in other parts of the sentence and serving different functions. For example:

35) 情报学 是 **第二次世界大战**

qing2bao4xue2　　shi4　　di4er4ci4shi4jie4da4zhan4

Information science　was　**the second world war**

后　　形成的　　新学科。

hou4　xing2cheng2de　xin1xue2ke1

after　establish　new Subject

Information science was a new Subject which was established after the second world war.

In this example, the word sequence 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the

second world war) functioned as part of a Complement; 70% of all instances (175 out of 250) belonged to this category. The second possible function was as an Object. The major semantic association to the left when 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) functioned as part of an Object was CREATION, including the semantic members 制定 *zhi4ding4* (draw up)*, 研发 yan2fa1* (to research and develop)*,* and 建立 *jian4li4* (to establish)), which differed from when the combination appeared at the initial position of a sentence. This finding indicates that semantic association with 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) might depend upon certain grammatical conditions being met.

7.9 Summary

In this chapter, we have tested the validity of Lexical Priming Theory being applied to Chinese to answer the first and most important question in the thesis—i.e., *Can collocation, colligation and semantic association as defined in Lexical Priming Theory account for the range of lexical and syntactical behaviour in Chinese.* Regarding collocation, as defined in the theory, 世界 *shi4jie4* co-occurred with particular characters or combinations of characters more often that can be attributed to random distribution. By comparing the collocates of 世界 *shi4jie4* with collocates with *world*, we found that a 'word' may need to be redefined for Chinese. Because of the two languages' different writing systems, a given word in English does not always corresponds to an individual character in Chinese. It is suggested that the definition of nesting in Lexical Priming Theory be applied when conducting

cross-linguistic studies between English and Chinese.

At the secondary stage, we studied the colligational behaviour of 世界 *shi4jie4*, finding it prefers/avoids a particular grammatical position and function. In addition, 世界 *shi4jie4*, both as an individual word or as a component of a word sequence, preferred a specific position, either occurring at the beginning or end of a sentence. This corresponds to colligation, as defined in LPT.

Thirdly, we studied the semantic associations of 世界 *shi4jie4* across grammatical functions. As with the English *world*, 世界 *shi4jie4* also showed a distinctive semantic associational profile at different grammatical positions while occurring within different grammatical functions. Compared with *world*, 世界 *shi4jie4* had a strong preference for participating in premodification in the form of noun phrases, including associative phrases, modifying phrases, subject-predicate phrases, and verb-object phrases. We also extended our examination to a further distance from 世界 *shi4jie4* and discovered textual semantic associations within more distant neighbours.

As a word is primed for collocational use, so too is a word sequence constructed of individual words. An investigation of the collocational and semantic associational profiles of nestings containing 世界 *shi4jie4* was therefore carried out, yielding two main results. First, nestings containing 世界 *shi4jie4* shared collocates and semantic associations with its components, as had been claimed about the English *world*. This is not a surprising finding, though we could use it to claim a modification to Hoey's

definition of nesting. The components making up a combination might be collocates of each other as well as collocates of the combination as a whole. What Hoey intended to emphasise in his definition was that nesting can be seen as a product of a priming of individual word that itself becomes primed in ways different from the components constituting it. The modification we intend to add, based on data found for *world* and 世界 *shi4jie4*, therefore, is that as a product of priming of individual words, nesting may share collocates and semantic associations with its components. However, the shared collocates and members forming the semantic set of a particular semantic association may differ in terms of their collocational strength, and the primings of a nesting sometimes do not apply to its constituent words.

In addition to the collocational findings for 世界 *shi4jie4*, we also found that 世界 *shi4jie4* and nestings containing 世界 *shi4jie4* may be primed to occur in a particular grammatical position or with a particular grammatical function. 世界 *shi4jie4* and the word sequence containing it were primed to favour or avoid a grammatical association. Extending Halliday's claims about colligational relationships in terms of sentential position, Hoey formulated his definition of colligation by proposing that a word or word sequence also has a preference/avoidance for a place in a sentence. Based on our examinations of 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) and 在世界范围内 *zai4shi4jie4fan4wei2nei* (in the range of the world), we found evidence to support LPT's claims about sentential preference/avoidance. We witnessed how 世界 *shi4jie4's* patterns of use were characteristically controlled by its colligation and in turn primed for particular semantic purposes through nestings.

So far, we have examined the collocational, colligational and semantic associational profiles of *world* and 世界 *shi4jie4*, as well as studying situations that occurred with complex issues. The next chapter compares the lexical and syntactical characteristics of the two node words, based on the findings from Chapters 4 to 7.

.

CHAPTER 8

Data Interpretation and Discussion

8.1 Introduction

A systematic investigation of the collocational, colligational and semantic associational features of the English *world* and the Chinese 世界 *shi4jie4* has been carried out in Chapters 4-7. Quantitative analysis was conducted by examining data retrieved from the two general corpora (the BNC and the zhTenTen11), supplemented by qualitative analysis of data selected from concordance lines containing the two nodes in the general corpora mentioned above and two comparable corpora (the FLOB and the LCMC). This chapter will discuss the findings detected from *world* and 世界 *shi4jie4* to provide informed answers to the research questions set in Chapter 1.

8.2 Collocational features of 世界 *shi4jie4*

Hoey illustrated in LPT that we hold in our minds elaborate networks of possible co-occurrence patterns that are primed for particular contexts or situations. Per the theory, "mental concordance" stored in our mind sparks the expectancies we refer to build up discourses. This process is called priming in LPT. In essence, the starting point of LPT is that through repeated encounters with a word combination or word sequence in a particular textual and social context, we begin to identify it as collocation. Repeated encounters with the collocation are then primed when we are

exposed to similar contexts. The starting point for of considering the applicability of LPT to Chinese, therefore, must be collocation. The natural phenomenon of collocation in Chinese does not need to be focused on here, as the existence of this linguistic feature has been demonstrated by linguists (e.g., Hoey and Shao 2015, Xiao and McEnery 2006) and in this study. What we seek here is to determine whether collocation, as defined in LPT, operates the same way for 世界 *shi4jie4* as for *world*. Our study has teased out statistical data to show some collocates occurred more frequently than others on both sides of 世界 *shi4jie4* and of nestings containing 世界 *shi4jie4*. Collocates are lexical items expressed either by a single character such as 全 *quan2* (all), 居 *ju1* (rank), and 占 *zhan4* (to take or to account for a certain proportion of), or a combination of two characters, such as 当今 *dang1jin1* (current) and 整个 *zheng3ge4* (whole). Looking at the tables of collocates presented in Chapter 6, we can find that the majority of items occurring as collocates to both sides of 世界 *shi4jie4* were combinations of two lexical items, as exemplified above. As specified in LPT, collocations occur in the form of recurrent word combinations. However, the notion of "word" itself needs to be re-defined before we can account for collocation in Chinese. We discussed in Chapter 1 that the definition of "word" in English is not tied to the Chinese writing system, which leads to confusion when determining basic units of "equivalence" across English and Chinese. Chinese is a language with a complex word structure. A lexical meaning expressed by a single item in English could correspond to a combination of more than one item in Chinese; for example, *buy* — 购买 *gou4mai3* and *bicycle* — 脚踏车 *jiao3ta4che1*. We propose here to use the concept of *nesting* in LPT to refer to a Chinese character combination that has a meaning equivalent to its English counterpart (word).

According to LPT theory, *nesting* is a property where "the product of a priming becomes itself primed in ways that do not apply to the individual words making up the combination" (Hoey 2005:8). However, data detected from both *world* and 世界 *shi4jie4* imply a modification to the second half of that definition, in that the collocation primed for a *nesting* can also apply to the components constituting that combination, though differentiating in terms of collocational strength as measured by statistical scores (e.g., LogDice). A Chinese nesting is made up of a sequence of characters with a strong collocational strength of co-occurring with each other. The nesting has a featured priming that cannot apply to its components (characters); however, it can share some primings (with a weaker collocational strength) with its constituents.

After proposing a definition of "word" for cross-linguistic study between English and Chinese, we revisit the issue of how collocation may operate for Chinese. Supported by data found for 世界 *shi4jie4*, we conclude that collocation in Chinese can occur between both characters and *nestings*.

An influential factor of LPT is the environment surrounding language users. The more primings the speaker encounters, the more possibilities to prime. If the concepts of priming and collocation in LPT are correct, it then follows that members of a community will, to a degree, influence (prime) each other through encounters within this community, and that these primings in turn will be stored in the heads of community members. This would mean that unique features of Chinese collocation ought to be found.

Within the speech community of Mandarin speakers in mainland China, there are particular combinations of lexical items (nestings) made up from 世界 *shi4jie4* that, though not unique, appear to be more strongly preferred than word clusters containing *world*; for example, 世界经济 *shi4jie4jing1ji4* (the world economy) (30,978 concordance lines), 成为世界上 *cheng2wei2shi4jie4shang4* (become … in the world) (17,355 concordance lines), 在世界范围内 *zai4shi4jie4fan4wei2nei* (in the range of the world) (13,591 concordance lines), 走向世界 *zou3xiang4shi4jie4* (to walk to the world) (13,380 concordance lines), and so forth. Nestings shared between the two languages differ markedly in their frequency of per million running words; for example, *the second world war* occurs 13.70 times per million running words in the BNC while 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) occurs 5.60 times per million running words in the zhTenTen11. With regard to the collocates occurring on both sides of the two nestings, we found, in Chapters 4 and 5, that collocates co-occurring with *the second world war* were associated with meanings of PARTICIPANTS OF THE WAR (e.g., Nazi(s), Germans) and TIME PERIOD (represented with prepositions such as *during*, *after* and by specific years, such as *1939*, *1940*), while collocates co-occurring with 第二次世界大战 *di4er4ci4shi4jie4da4zhan4* (the second world war) can be grouped into a semantic set associated with a meaning of CREATION (represented with 制定 *zhi4ding4* (draw up), 研发 *yan2fa1* (to research and develop), and 建立 *jian4li4* (to establish)).

Above, we pointed out how collocation, as defined in LPT, occurred for the Chinese node 世界 *shi4jie4*. At this stage, we have only looked at one of the analysis units at

the lexical level. We discussed, in Chapter 2, that a comprehensive study shall be undertaken by combining multi-layered factors. In the next section, therefore, we will discuss how colligations work for 世界 *shi4jie4* to provide a holistic analysis of the node and its counterpart.

8.3 Colligational features of 世界 *shi4jie4*

Before discussing the colligational features of 世界 *shi4jie4*, the definition of colligation in LPT is revisited here to recall the readers' memory. Hoey (2005:43) suggested that colligation can be defined as:

> 1. the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank;
> 2. the grammatical functions preferred or avoided by the group in which the word or word sequence participates;
> 3. the place in a sequence that a word or word sequence prefers (or avoids).

As discussed in Chapter 2, Hoey's definition of colligation does not overturn previous ones; rather, it was suggested and initiated by many other linguists, such as Halliday (1957), Sinclair (1991) and Stubbs (1996). What makes Hoey's formulation of colligation distinctive is in his claim that a colligational statement can be negative as well as positive in terms of the group in which it occurs, the grammatical functions within which it serves, or its place in a sentence that it posits. We showed, in Chapter 2, that no published research (to the author's knowledge and up to the completion of the present project) has investigated colligational behaviour for Chinese as Hoey has done for English. A holistic picture needs to be drawn to provide ample evidence for the existence of colligation (as defined in LPT) in Chinese.

The categorisation of colligation in LPT mainly concerns that colligational statements can be either negative or positive. As noted in Chapter 1, that words can express a meaning of preference/avoidance in this thesis does not mean lexical items have "emotions." The terminologies used in the project are only a shorthand for describing the user's psychological tendency to choose a word or word sequence at a particular position within a particular grammatical function. Although the psychological aspect of this claim cannot be demonstrated by corpus data, we are able to demonstrate particular grammatical patterns in a particular piece of language. The frequency of occurrences within a grammatical company, grammatical function, or at a place in a sentence imply a consequent availability for being primed in that language. The remainder of this chapter seeks to show how a colligational description of 世界 *shi4jie4* might proceed and how its colligational behaviour ties in with the concept in LPT. Finally, it will seek to use LPT to account for the colligational similarities and differences found for *world* and 世界 *shi4jie4*.

8.3.1 The existence of colligation of 世界 *shi4jie4*

We have shown, in Chapter 7, that 世界 *shi4jie4* has a strong preference for occurring with PRESENT tense. While the normalised PRESENT tense distribution in the LCMC is 67%, we found 94% of instances containing this node word preferring to occur in a context containing PRESENT tense. The significantly strong preference for PRESENT tense in turn implies a correspondently negative colligational association with the other two tenses: PAST and FUTURE. The PAST tense distribution of the LCMC is 32%; however, an examination of the instances containing 世界 *shi4jie4* indicates that only 5% instances prefer to occur within this

tense.

Another colligational feature we investigated concerns the grammatical functions of 世界 *shi4jie4* in a clause. 世界 *shi4jie4's* repulsion of or aversion to a particular grammatical function was determined by comparing it with five comparators—结果 *jie2guo3* (result), 问题 *wen4ti2* (problem), 语言 *yu3yan2* (language), 工业 *gong1ye4* (industry) and 范围 *fan4wei2* (range). We found that 58% of instances extracted from the LCMC had a clear positive colligation of occurring as part of an Object, which was almost three times more often than the comparator 语言 *yu3yan2* (language). 世界 *shi4jie4* also had a preference for occurring as part of an Adjunct, though not as strong a preference as for occurring within the Object function. On the other hand, the node word had a negative colligation with the Complement and Subject functions, ranging from one-sixth to one-third the frequency of the comparators.

With regard to grammatical position, there were also positive and negative colligations between 世界 *shi4jie4* and its position within a grammatical category. In 66% of the instances containing 世界 *shi4jie4*, it occurred as a part of a premodification, particularly within an Object or Complement, which is two to three times more often than the other nouns occurring at this position. On the other hand, 世界 *shi4jie4* had a clear negative colligation for occurring as a noun head, especially when serving as part of an Object, which is almost three times less often than the other nouns. These colligational preferences (or avoidances) help generalise the grammatical patterns in which 世界 *shi4jie4* preferred to occur:

Verb + 世界 *shi4jie4* premodifies NP

BE + 世界 *shi4jie4* premodifies NP

We discussed, in Chapter 2, that previous cross-linguistic studies between English and Chinese focused on lexical comparison and contrast, such as collocation and semantic prosody. Few studies have done grammatical comparisons between the two languages, perhaps due to the traditional view that Chinese is a semantically connected language while English is systematically linked language. From the data showing the existence of colligation, as defined in LPT, for 世界 *shi4jie4*, we can see the potential feasibility of describing Chinese grammar with categories notified in the theory.

8.3.2 Similarities and differences of colligational behaviour between *world* and 世界 *shi4jie4*

We have shown that Chinese, at least witnessed by instances containing the node word 世界 *shi4jie4*, has colligational preferences for and aversions to occurring within different grammatical functions. To determine whether 世界 *shi4jie4's* preferences or aversions are in line with those of its English counterpart, *world*, we need to recall data detected in Chapters 4-7 to make a holistic comparison.

8.3.2.1 Similarities of colligational behaviour between *world* and 世界 *shi4jie4*

First, the nodes *world* and 世界 *shi4jie4* both had a preference for being used with the PRESENT tense, but the latter's preference was of higher significance. Second, *world* and 世界 *shi4jie4* both preferred to occur within the Object function, but avoided occurring within the Complement function.

Regarding their position in a nominal group, the two node words had a clear negative colligation with being a noun head in a nominal group. Coincidentally, in 33% of instances, both *world* (119 out of 365) and 世界 *shi4jie4* (508 out of 1535) occurred as a head of a nominal group, which was in both cases a far lower proportion than the comparison words.

However, *world* and 世界 *shi4jie4* had a strong preference for occurring as part of a premodification preceding a noun head. It was not surprising to discover that 世界 *shi4jie4* had a preference for occurring as part of a premodification of a nominal group, because Chinese is a language that tends to have more premodifications than postmodifications, perhaps because Chinese does not have various prepositional phrases or phrases led by *of*. What was surprising was that *world* also had a strong preference for occurring at this position when compared with its five comparators—*consequence*, *question*, *preference*, *aversion* and *use*. The proportion of instances containing *world* occurring as part of a premodification was 25%, four times that of *aversion* (6%) and 417 times that of *consequence* (0.06%). When being part of a premodification, both *world* and 世界 *shi4jie4* preferred to appear in the form of individual words to modify subsequent words, and were relatively unlikely to occur in their possessive form — *world's* and 世界的 *shi4jie4de* (world's).

8.3.2.2 Differences of colligational behaviour between *world* and 世界 *shi4jie4*

Since English and Chinese are two distinct languages in many aspects, it is not surprising that their colligational differences are greater than their similarities. However, these differences deserve closer attention and detailed interpretation to explore where English and Chinese differ, and whether Lexical Priming Theory can be used to account for these differences.

It has been noted that both *world* and 世界 *shi4jie4* had a clear preference for occurring within the Object function, but differed in their grammatical positions in a nominal group when doing so. *World* had an equal preference for occurring as both part of a postmodification and a head (34%, 41 out of 120) of a nominal group in an Object, while 世界 *shi4jie4* had a clear preference for being used as part of a premodification (75%, 200 out of 268).

In relation to their occurrences in other grammatical functions *world* and 世界 *shi4jie4* differed significantly. While there was no interesting colligational relation between *world* and the Subject function (see Table 4.6), there was a clear negative colligational relation between 世界 *shi4jie4* and the Subject function (see Table 7.2), where only the comparison word 范围 *fan4wei2* (range) came close to the distribution found for 世界 *shi4jie4*. To compensate, there was a positive colligation between 世界 *shi4jie4* and the Complement function, which was six times less frequent than for 语言 *yu3yan2* (language) and 1.3 times less often than for 结果 *jie2guo3* (result). On the other hand, there was a positive colligation between *world*

and the Complement function — though not as strong as for the Object function. What might not be in line with our expectations is that there was a negative colligation between *world* and the Adjunct function; however, our findings show a clear positive colligation between 世界 *shi4jie4* and the Adjunct function. Our intuition (as a non-native speaker) or language wisdom (or what the Prince of Grammar called *yu3gan3*) leads us to assume that *world* might have more uses as an Adjunct because English seems to have more prepositional phrases than Chinese, and one of the distinctive uses for a prepositional phrase is to function as a part of Adjunct. Since there was a number of occurrences of *in the world* in our data, it is reasonable to hypothesise that *world* might function more frequently as an Adjunct than in other functions. However, our data revealed the opposite. *World* was used less frequently within the Adjunct function than the five comparison words, occurring within an Adjunct in less than 20% of cases, while other words occurred within this function between 25% and 50% of the time. These findings remind us that one should never "take linguistic phenomenon for granted in the armchair," but should go to the text and see what it tells us.

Examination of instances containing *world* and 世界 *shi4jie4* reveals complex issues concerning its grammatical position in each function. In general, the divergence lies in the convergence. When functioning as a Subject, 世界 *shi4jie4* preferred to occur as a noun head in a nominal group (see Table 7.4), whereas *world* preferred to occur as part of a postmodification. When functioning as a Complement, 世界 *shi4jie4* had a strong preference for occurring as part of a premodification (75%), while *world* preferred to occur as part of a postmodification.

The similarities and differences demonstrated for *world* and 世界 *shi4jie4* have so far been discussed in terms of general collocational and colligational profiles. Since we have emphasized several times in the thesis that our goal is to provide a holistic picture for the two nodes examined in the thesis, we must bring more factors into the discussion to provide a comprehensive picture. The literature review demonstrated that previous cross-linguistic studies were conducted on one or two contextual levels, which perhaps sometimes could not provide a holistic comparison of the languages. Thus, in the following sections, we will discuss the similarities and differences derived from complex situations involving *world* and 世界 *shi4jie4* to fill the gap by combining more units; i.e., collocations, colligations and semantic associations.

8.4 Comparison between *world* and 世界 *shi4jie4* in complex situations

The comparison will be carried out by considering the grammatical positions of *world* and 世界 *shi4jie4* within the Subject, Object, Complement and Adjunct functions. Collocational and semantic associational features at positions of noun head, premodification and postmodification will be taken into account. The convention used is that a table summarising the general positive or negative colligational feature will be drawn at the beginning of each section, followed by a detailed description of the comparison. The convention for presenting the general colligational profiles is that "+" refers to a positive colligation, while "–" refers to a negative colligation and "×" means no significant preference/aversion was found.

8.4.1 Comparison within Subject

Table 8.1 A grammatical distribution of *world* and 世界 *shi4jie4* within Subject

	Subject	
	world	世界 *shi4jie4*
noun head	×	+
part of a premodification	−	+
part of a postmodification	+	−

At the position of noun head

In Chapter 5, we discovered that the verbs following *world* when it appeared in a noun head position could be grouped into a semantic association of CHANGE, represented by verbs such as *change*, *turn*, *shrink* and so on. Another distinctive feature of *world* when appearing as noun head was its highly frequent occurrence with BE verbs. After excluding instances where the BE group served as grammatical verbs, the remaining instances (495) included 440 instances of the structure *world* + BE + adjective, and only 55 instances of the structure *world* + BE + noun. Moreover, the adjectives occurring in the former were words expressing a pessimistic point of view, such as *vulnerable* and *dim*.

When we examined R1 verbs in cases where 世界 *shi4jie4* appeared as noun head within a Subject, we found the lexical item 变成 *bian4cheng2* (become) occurred with the highest LogDice score (5.714), similar to BE's score for co-occurring with *world* as a Subject (5.762). The similar collocational strength exhibited between *world* and BE and between 世界 *shi4jie4* and 变成 *bian4cheng2* (become) suggests that BE and 变成 *bian4cheng* (become) showed a similar preference for the two

nodes co-occurring with the two verbs when functioning as a Subject. There were 230 instances of 世界 *shi4jie4* + 变成 *bian4cheng2* (become) found in the zhTenTen11, of which 138 contained 世界 *shi4jie4* as a noun head within a Subject. Similarly, in the structure *world* + BE, words collocating with the combination 世界 *shi4jie4* + 变成 *bian4cheng2* (become) also consisted of two parts of speech: adjectives and nouns. The adjectives were associated with a meaning of HAPPINESS, including words such as 美好 *mei3hao3* (glorious) and 幸福 *xing4fu2* (happy), while the nouns were associated with a meaning of PLACE, such as 地球村 *di4qiu2cun1* (global village) and 地球教室 *di4qiu2jiao4shi4* (global classroom).

The above data show that when occurring within a Subject as a noun head, both *world* and 世界 *shi4jie4* tended to be followed with BE words to form a grammatical pattern of "world / 世界 *shi4jie4* + BE + ADJ." As opposed to the former, the latter group of speakers tended to describe the planet on which they were living in a positive way.

At the position of a postmodification

We have revealed that *world* occurred at the position of a postmodification within a Subject in the form of the phrase, *in the world*. The noun heads postmodified could be grouped into a semantic association of ORGANISATION. Since *in the world* had a strong colligation with SUPERLATIVE words, the semantic sets belonging to ORGANISATION had a positive preference for collocating with superlatives, such as *the oldest*, *the largest*. However, no instance occurring at this position was found

for the Chinese 世界 *shi4jie4*.

At the position of a premodification

Chapter 5 also showed that only 17% of instances in the sample database contained *world* appearing as part of a premodification within a Subject; within these instances, *world* occurred as an individual word and in the form of *world's* almost equally (51% vs. 49%). While words modified by *world* were hard to group into any semantic association, words modified by *world's* tended to be associated with PROBLEMs encountered by the world.

In comparison, 世界 *shi4jie4* had a strong preference for appearing as part of a premodification within every grammatical function. When serving as a Subject, 世界 *shi4jie4* premodified verbs, and the combination containing 世界 *shi4jie4* + VERB in turn formed a premodification for the noun head of the Subject. The verbs collocating with 世界 *shi4jie4* in this function tended to be associated with meanings of ranking in leading place, such as 领先 *ling3xian2* (to pioneer), 首创 *shou3chuang4* (to create something first time in the world), 惊叹 *jing1tan4* (to amaze) and 首屈一指 *shou3u1yi4zhi3* (to be the first and foremost).

It was noted in Chapter 4 that there were no striking differences between *world* and the other five comparison words within the Subject function (see Table 4.6 for detailed distributions). However, a quite striking difference was shown by 世界 *shi4jie4* within a Subject when compared with 结果 *jie2guo3* (result), 问题 *wen4ti2*

(question), 语言 *yu3yan2* (language), 工业 *gong1ye4* (industry), and 范围 *fan4wei2* (range). There was a clear negative colligation between 世界 *shi4jie4* and the grammatical function of Subject. The difference may imply that Chinese speakers are unlikely to thematise the node word 世界 *shi4jie4* and that the thematisation of *world* is not as significant as for 世界 *shi4jie4*.

Using SPSS, we found a significant difference between *world* and 世界 *shi4jie4* occurring within the Subject function ($p < 0.015$, sig.=0.000), indicating a more positive colligation between *world* and the function of Subject when compared with its Chinese equivalent. This may imply that English speakers are likely to define or justify the *world* surrounding them, an implication supported by their high frequency use of the structure the *world* + BE; however, Chinese speakers are unlikely to do so. Within the relatively fewer uses of 世界 *shi4jie4* within this grammatical function, 世界 *shi4jie4* tended to PREMODIFY verbs expressing "polar" meanings, such as 领先 *ling3xian2* (to pioneer) and 首屈一指 *shou3u1yi4zhi3* (to be the first and foremost).

8.4.2 Comparison within Object

Table 8.2 A grammatical distribution of *world* and 世界 *shi4jie4* within Object

	Object	
	world	世界 *shi4jie4*
noun head	×	−
part of a premodification	×	+
part of a postmodification	×	−

Table 4.6 shows *world* did not demonstrate a significant preference for or aversion to appearing within an Object (sig.=0.181). In comparison, 世界 *shi4jie4* shows a strong colligational preference for appearing within Object function (sig.=0.032). In addition, 世界 *shi4jie4* demonstrated a more positive colligation with premodification (sig.=0.006) than the other abstract nouns selected in the present study. However, there was no significant preference for or avoidance of *world* occurring at any grammatical position within an Object.

At the position of noun head

Verbs preceding *world* when appearing as noun head within an Object were grouped into three semantic associations—SHAPE, GOVERN, and to KNOW THE WORLD—whereas verbs preceding 世界 *shi4jie4* in cases where it appeared as noun head within an Object were associated with meanings of ENTER, CONSTRUCTION, RANK and ENJOYING FAME.

There were no semantic associations shared by *world* and 世界 *shi4jie4* at the grammatical position of noun head within an Object; however, they shared collocations of *change the world* (8.07) and 改变世界 *gai3bian4shi4jie4* (to change the world) (8.149); *to understand the world* (6.79) and 认识世界 *ren4shi1shi4jie4* (to know/understand the world) (6.40); *enter into the world* (8.07) and 走进世界 *zou3jin4shi4jie4* (to enter into the world) (8.187). These similar collocations all occurred with similarly high LogDice scores, revealing that both English and Chinese speakers use these three collocations frequently. English and Chinese people

tend to explore the planet on which they are living and are likely to make changes to have a better life or establish a better relationship between human beings and nature.

At the position of a premodification

世界 *shi4jie4* showed a high significance of occurring in the grammatical position of a premodification within an Object (sig.=0.006). The top clusters containing 世界 *shi4jie4* as a modifier within an Object were 世界第一 *shi4jie4di4yi1* (world first) (8.55), 世界一流 *shi4jie4yi1liu2* (world best) (7.25), 世界先进 *shi4jie4xian1jin4* (world advanced) (6.08), 世界各地 *shi4jie4ge4di4* (everywhere in the world) (7.87), and 世界上 *shi4jie4shang4* (7.57). Verbs collocating with 世界第一 *shi4jie4di4yi1* (world first) were 居 *ju1* (to rank), 跻身 *ji1shen1* (to make one's way to [a higher place]), and 跃居 *yue4ju1* (leaps to first place). These words belonged to a semantic set forming the semantic association of RANK.

English and Chinese speakers recognise the polar meanings of *world* and 世界 *shi4jie4*; however, Chinese people are more likely to apply 世界 *shi4jie4* to modify words expressing RANKING meaning. With this usage, *world's* top place in a range is reinforced and highly respected.

Where *world* appeared as a premodification, the most frequent collocations were *world community*, *world record*, *world champion* and *world title*. It can be noted from these collocations that *world* in English is used to emphasise the top place in a particular field; however, words collocating with *world* in these cases were unlikely

to have a significant polar meaning in the way of words collocating with 世界 *shi4jie4* in Chinese.

We have so far discussed the differences in *world* and 世界 *shi4jie4* in terms of their colligational behaviours at different positions within the grammatical Object function. It appears there was a consistently high degree of divergence between *world* and 世界 *shi4jie4* in terms of their collocational and semantic associational features.

8.4.3 Comparison within Complement

Table 8.3 A grammatical distribution of *world* and 世界 *shi4jie4* within Complement

	Complement	
	world	世界 *shi4jie4*
noun head	×	−
part of a premodification	+	−
part of a postmodification	×	×

Data collected in Chapter 5 revealed that *world* had a clear negative colligation for occurring within a Complement. With regard to grammatical position, *world* preferred appearing at the position of a premodification in its possessive form *world's*. The ratio between *world* occurring as an individual modifier and occurring in possessive form was 3:1 within a Complement. Words premodified by *world's* then formed a colligational group of SUPERLATIVES, with both negative and positive meanings.

World's Chinese corresponding equivalent 世界 *shi4jie4* also showed an aversion to appearing within a Complement, appearing in only 37 of 535 instances within this function. The most frequent combination 世界 *shi4jie4* tended to form as a premodification was 世界上 *shi4jie4shang4* (in the world) (54%), the second most frequent combination (28%) was 世界 *shi4jie4* + Noun + 最 *zui4*. As noted in Chapter 6, the character 最 *zui4* in Chinese is equivalent to Superlatives in English when followed by adjectives such as 最大 *zui4da4* (the biggest/largest), 最好 *zui4hao3* (the best), and so forth. Thus, in this sense, *world* and 世界 *shi4jie4* could be seen as sharing the same colligational (and semantic associational) feature of co-occurring with Superlatives, even if this similarity is manifested by different collocates.

Apart from these two kinds of collocations, 世界 *shi4jie4* also co-occurred with words expressing polar meanings, such as 著名 *zhu4ming2* (famous), 顶尖 *ding3jian1* (top) and 第一 *di4yi1* (first), and words expressing PLACE, such as 各 国 *ge4guo2* (every country) and 各地 *ge4di4* (everywhere), to form a combination and in turn serves as a premodification within a Complement. What distinguishes 世 界 *shi4jie4* from *world* at the position of premodification is that the former was likely to collocate with words associated with a meaning of PLACE to form a modifying phrase that further premodified the noun head following it.

8.4.4 Comparison within Adjunct

Table 8.4 A grammatical distribution of *world* and 世界 *shi4jie4* within Adjunct

	Adjunct	
	world	世界 *shi4jie4*
noun head	+	−
part of a premodification	−	+
part of a postmodification	×	−

Chapter 4 and Chapter 6 showed that, while *world* had a significant aversion for occurring within an Adjunct (sig.=0.015), 世界 *shi4jie4* did not show a clear preference for or avoidance of this grammatical function (p＞0.05) compared with the five comparators.

The dominant combination containing *world* within the Adjunct function was (*first/second) world war*. On the other hand, no instances of 第一/二次世界大战 (first/second world war) serving as an Adjunct were detected from the LCMC. Among the remaining instances occurring within this grammatical function, 60% instances contained *world* appearing as a noun head (*in the world (of…)*) and 40% contained *world* appearing as part of a postmodification (*…of the world*).

Although 世界 *shi4jie4* did not have a clear preference for or aversion to occurring within this grammatical function, it had a strong preference for occurring in two combinations when functioning as an Adjunct, namely 世界上 *shi4jie4shang4* and (在)世界范围内 *zai4shi4jie4fan4wei2nei4* (in the range of the world). In Section 7.3, we mentioned that the two combinations 世界上 *shi4jie4shang4* and (在)世界范围内 *zai4shi4jie4fan4wei2nei4* had higher MC values (32% and 49%, respectively) than the English prepositional phrase *in the world*. Many such combinations occurred in

the BNC and the zhTenTen11; thus, the differences between these combinations in the two languages deserve our close attention.

The 365 sentences extracted from the FLOB included 190 instances of *in the world*. Of these, 36% (69) occurred in the position of a postmodification in the form of *in the world*, and 23% (44) in the position of a postmodification, with *world* being postmodified by following *of* (i.e., *in the world of…*) within the functions of Subject, Object and Complement. The remaining 51 instances occurred within the Adjunct function, excluding 26 instances that were cases of benign ambiguity, meaning they could be considered Adjuncts or postmodifications (specified in Section 4.4.2).

Turning to the two Chinese equivalent versions of *in the world*, 世界上 *shi4jie4shang4* avoided appearing within the functions of Subject and Adjunct, but preferred to occurred within the Complement function. It was typically used to colligate positively with being part of a premodification; 世界 *shi4jie4* in this combination had an aversion to being either pre- or post-modified.

By contrast, the combination *(在)世界范围内 zai4shi4jie4fan4wei2nei4* had a strong preference for occurring within the Adjunct function, but a strong aversion to occurring within a Complement (20 of 1000 instances). However, it had a strong tendency to premodify the noun head within the Subject and Object functions.

A comparison of *in the world* and its two equivalent Chinese combinations, whose MC values were relatively high, reveals that these two counterparts not only differed in their colligational behaviours cross-linguistically, but also in terms of preference

and avoidance for a particular grammatical position within a grammatical function.

8.5 A discussion of the comparison

While collocation in Chinese has been studied plentifully by cross-linguists, few studies have investigated of its colligation. The project has conducted a comprehensive examination of the colligational behaviour of a Chinese node word (世界 *shi4jie4*) and the nestings containing of it. Applying the notion of colligation as defined in LPT, 世界 *shi4jie4* was investigated at multiple levels. The data analysis was organised into three phases. In the first phase, we analysed the positions of 世界 *shi4jie4* in a nominal group—namely as a head, as part of a premodification, and as part of a postmodification. In the second phase, we described the grammatical behaviour of 世界 *shi4jie4* within a higher rank — i.e., the grammatical function served. In the third phase, we moved from standard colligation analysis to co-colligational analysis; i.e., we described the co-relationship across the lower and higher group by taking their collocation and semantic association into account. To the best of our knowledge, exploring colligation from this perspective has not been addressed before in the literature applying LPT for either English or Chinese. We have shown, using evidence from corpora, that 世界 *shi4jie4* and *world* both had positive/negative colligations with a particular position when serving a particular grammatical function. The phenomenon can be generalised by revisiting the following table, originally found in Chapter 7:

	premodification	Noun head
Subject	×	—
an Object	+	—
Complement	+	×
Adjunct	+	—

The table reveals an interaction between 世界 *shi4jie4* occurring at a lower rank of grammatical position (inter-colligation) and occurring within a higher rank of grammatical function (intra-colligation), which shows that colligation operates for Chinese both inter-colligationally and intra-colligationally. As reviewed in Chapter 2, previous studies of LPT focused on a single unit of analysis, such as collocation or colligation. Our data show there is an interaction operating at more than one level. By combining the examination of collocation and semantic association, we found indications that particular colligations exerted on collocations were related to the semantic associational properties of such collocations. The conclusion drawn from examining the colligations of both *world* and 世界 *shi4jie4* adds a further, fine-grained parameter for describing the colligational behaviours of a language; specifically, that colligation is primed in an interaction with collocation and semantic association and the parameters interchange with and are dependent on each other.

Chapter 2 of this thesis discussed the views held by linguists on the relationships between lexis, grammar and meaning. The extreme version of the traditional view holds that grammar is generated first and words are dropped into the grammatical construction thereby created, or that the semantic is generated first and lexical items merely fall as members of the semantic set. Supported with corpus data, corpus

linguists see lexis as being as important as grammar. The interaction between collocation, colligation and semantic association operating on 世界 *shi4jie4* implies that the relationship between grammar and lexis in Chinese may need to be re-considered as well. Compared with English, which has systematic description of grammar, the grammar system of Mandarin Chinese is rather loose. The formulation of Modern Chinese grammar originates from the 1898 book, *Ma Shi Wen Tong* (*Ma's Grammar Book*), which is viewed as a symbol of the birth of Modern Chinese grammar. Ma claimed, in his book, that Chinese is a paratactic language that focuses on the expression of meaning, while not paying close attention to such formalities as word order. According to him, infinite meaning can be expressed with finite Chinese characters[26]. Ma described modern Chinese as a character-based language, meaning it could be expressed by infinite changes of word order in a sentence or within a cluster. While SVO structures are awarded supreme priority in English, with other elements serving as ancillaries, Chinese occupies a linear sentence structure in which the topic claims the head position in a sentence, followed by a series of comments to develop the topic. The topic and comments may be linked loosely from the perspective of grammar, but semantic relevance is connected closely in Chinese sentences. Chen (1998:30), echoing Ma's claims about the features of Chinese, posited that Chinese sentences are formed around the "thought-pivot" structure, while English sentences are formed around the "form-pivot or SV-pivot" structure. The former focuses on the semantic meaning of a sentence's structure rather than its completeness, while the latter stresses the formal completeness of a sentence.

Based on results obtained from this project, a dependency correlation exists across

---

[26]助字者，华文所独，所以济夫动字不变之穷。中国文字无变也，所以介字济其穷。

collocation, colligation and semantic association for 世界 *shi4jie4* and its nestings. We extended our examination to a textual relationship, as defined in LPT (textual association and colligation), and found that nested combinations containing the node were primed to occurred in specific types of semantic relationship; for example,当今 世界 *dang1jin1shi4jie4* (today's world) was positively primed in a COMPARISON relationship. The data therefore imply that paratactic Chinese can also be analysed from a hypotactic perspective, depending on how one defines grammar and sees the relationship between lexis and grammar in Chinese. Based on the data in this thesis, we suggest that grammatical functions in Chinese might be reformulated in terms of grammatical categories. LPT argues that the grammatical category assigned to a word is "simply a convenient label we give to the combination (some of) the word's most characteristic and genre-independent primings" (Hoey 2005: 154). That is, it is the outcome of a combination of factors, not the starting point for linguistic description. The grammatical category a word belongs to is its grammatical priming. We discussed at the beginning of this chapter that the definition of "word" in English is not suitable for Chinese, due to the distinctive writing system of Chinese. Here, we propose that the grammatical category to which a Chinese character and nesting is assigned will be an outcome of the character's or nesting's grammatical primings. The grammar of Chinese is an accumulation and interweaving of the primings of a character's or nesting's collocations, colligations and semantic associations. Components of nestings are primed to combine in a certain way (order), to occur at a certain position within a nominal group, and to serve a certain function in a sentence.

There is no literature (to our best knowledge) investigating the relationship between Chinese lexical items and grammar. In this sense, this project may show the potential

value of using LPT as a theoretical base for describing Chinese grammar and the correlation between Chinese lexis, grammar and meaning.

After examining the combinatory profile of collocation, colligation and semantic association for both *world* and *世界 shi4jie4*, we moved on to compare the colligational behaviours of *world* and *世界 shi4jie4* when appearing at different grammatical positions and serving different grammatical functions. Collocations and semantic associations at each position within the four grammatical functions were also investigated. We found, in early comparisons, that the lexical and grammatical differences between *world* and *世界 shi4jie4* were far more numerous than the similarities. However, when we examined the data as a whole, instead of restricting our discussion to particular collocations and semantic associations appearing at the same grammatical pattern, we found that the divergence occurs within the convergence. When frequent collocation and colligational behaviours are recorded, convergence can be found in particular frequencies of use and characteristic associations. Both English and Chinese speakers have a strong preference for applying such word sequences as *in the word/在世界上 zai4shi4jie4shang4, to change the world/改造世界 gai3zao4shi4jie4, the second world war/第二次世界大战 di4er4ci4shi4jie4da4zhan4*, and so forth. The divergence lies in the grammatical patterns in which these word clusters are primed; within the same grammatical pattern, we find different word choices are made. For example, *world* and *世界 shi4jie4* appeared as noun head within the same grammatical pattern Subject + Be + Adj, but different adjectives were chosen by speakers of the two languages.

Our data show no unique word combination containing *world* and *世界 shi4jie4* occurring for either English or Chinese but do show a high or low frequency of use for a particular nesting. With regard to semantic association, the data reveal that shared semantic association is primed by both English and Chinese speakers, but different word choices are made. As discussed many times in this thesis, LPT takes the word as the starting point for elaborating the whole theory. Thus, the different word choices made by English and Chinese speakers results in their unique colligational primings. For example, both *world* and *世界 shi4jie4* could be associated with a meaning of RANK. The former preferred to appear as part of a postmodification within an Adjunct, while the latter preferred to participate in a premodification within an Object. The phenomenon of shared semantic association occurring within different grammatical patterns or of shared semantic association constituted by different semantic sets between *world* and *世界 shi4jie4* could be found elsewhere within other grammatical functions.

Previous cross-linguistic studies between English and Chinese, as discussed in the literature review, have three features in common. First, cross-linguistic equivalence is decided according to the researcher/translator's personal language experience or with reference to existing bilingual dictionaries. Second, of the few cross-linguistic studies of English and Chinese that make use of LPT, few attempt a colligational comparison of the two languages. Third, most studies of this kind tend to study nodes at one or two contextual levels (e.g., collocation, semantic prosody), while not paying enough attention to providing a holistic picture at co-textual levels.

The first issue was dealt with by Wei and Li (2014), who adopted a reference

criterion developed based on Altenberg's Mutual Correspondence (MC value) as an indication of the translatability of comparators of two languages. Wei and Li selected equivalence according to the degree of correspondence, thus reducing the researcher's subjectivity when deciding on equivalent words. The second and third issue is addressed in the present project by using MC value as a tool to decide the initial equivalence between English and Chinese, and then further examining the nodes across the two languages at multiple levels. Data summarised earlier in this chapter indicate there was both divergence and congruence between *world* and *世界 shi4jie4* at all levels. Analysis of data at one level and across levels showed a dependency correlation among the three concepts examined in the study. In light of the kind of evidence presented in Chapters 6 and 7, the notion of priming and the operation of nesting could account, in a systematic way, for the move from a fundamental level (collocation, colligation) to the wider text level (textual semantic association, textual colligation) or even a more abstract level (semantic association). Thus, it is too early to close off the upper boundary for absolute "similarities" and "differences" in a contrastive study of two languages. Contextual interaction needs to be considered when undertaking a cross-linguistic study. For two genetically different languages, such as English and Chinese, restricting the comparison to only one unit of analysis may block us from getting a more comprehensive profile of both languages. This comparison could be extended from the lexical level to the grammatical level. LPT offers a potential perspective from which English and Chinese could be compared in terms of colligation, particularly the tendency of a node to occur within a group, a grammatical function, or a place in a sentence or text.

In addition to the new comparison perspective initiated by LPT, we also claim that

the study of Chinese grammar and the formulation of modern Chinese grammar ought to be done through comparison with other languages, such as English, to detect whether the grammatical categories applied for a particular language can be valid for Chinese and whether the language theory driven by other languages can shed new light for the linguistic description of Mandarin Chinese.

8.6 A Brief Conclusion

In this chapter, we have discussed data found in previous chapters from three perspectives. First, we discussed the feasibility of using LPT to account for the collocational features of 世界 *shi4jie4*. Analyses of collocates co-occurring with 世界 *shi4jie4* on both sides imply that collocation in Chinese can be primed in terms of both a single lexical item (character) or a nesting (a combination of characters). LPT claims priming starts with the word; however, the problem with 'word' remains, as Hoey (2005:158) realised in his book, that it is often "subsumed" within larger entities. There are many languages in which word boundaries are problematic, including Mandarin Chinese. Priming sometimes operates for Chinese at a unit level equal to the English "word" (in meaning), but more commonly larger (e.g., nesting). We suggest that setting a nesting as a basic equivalence pattern between English and Chinese would be a convenient starting point for cross-linguistic studies.

Second, we summarised data gained from the colligational examination of 世界 *shi4jie4* and discovered the node (and nestings containing the node) had colligational preferences for and aversions to occurring at different grammatical positions within different grammatical functions. In each case, the semantic associations also differed.

Examination of the semantic associational behaviour of 世界 *shi4jie4* indicated its semantic associations were strongly grammar restricted. The node or sequences containing the node co-occurred with semantic sets forming different semantic associations within different grammatical functions. Grammar, in this sense, can be seen as the generalisation of a word's collocational, colligational and semantic associational behaviours, rather than merely its grammatical categories and functions. This might enlighten Chinese grammarians to generalise Chinese grammar from a new perspective; i.e., counting grammar as the sum and accumulation of the primings of the most common sounds, syllables, collocations, colligations, and semantic associations. Grammar, thus, should be the outcome or product of individuals' lifelong primings within a language community. Primings bridge the relationship between lexis and grammar. As we collect and associate collocational priming by encountering them on our environment, we unconsciously sum up semantic associations and colligations, which in turn nest and combine with words associated with other semantics or colligations, thus giving rise to an incomplete, inconsistent and non-comprehensive (but nevertheless workable) grammatical system. This fragmented systems are then generalised as the grammar of a language.

Last, we compared the characteristic features of *world* and 世界 *shi4jie4* within grammatical functions, and found their divergence lay in their convergence. While *world* and 世界 *shi4jie4* were found within the same grammatical functions, they differed in terms of their occurrences at positions with the grammatical group. Although sometimes following a similar nest of colligation (e.g., Subject + Be +Adj), speakers of the two languages are likely to prime divergent word choices to fill in the category of Adj. A dependency correlation across collocation, colligation and

semantic association was witnessed for the two nodes and for nestings containing the nodes.

Comparing colligational behaviour across English and Chinese has rarely been done by linguists (with the exception of Shao (2017)). Even fewer studies have sought to compare the two languages across co-textual levels. Our data, derived by adopting concepts from LPT, reveal that a dependency interaction operates for both *world* and *世界 shi4jie4*, suggesting previous cross-linguistic comparisons between the two languages may have oversimplified their similarities and differences. Indeed, an interpretation of convergence or divergence depends strongly on the unit within which one studies and numbers of contextual features one takes into consideration. Cross-linguistic studies of English and Chinese could be undertaken, we suggest, from two perspectives. We can restrict the examination to one level of grammatical pattern—for example, within a nominal group where the node positions at a noun head (if it is a noun)—to determine the primings associated with that node. Alternatively, we can observe their primings in combination by looking at all the primings that contribute to the features of the node.

Chapter 9

Conclusion

9.1 Introduction

The thesis has been concerned with two main issues. The first and major one was whether Lexical Priming Theory, which has been applied to English from many linguistic perspectives, could also be applied to Chinese. The second issue was how LPT can be advanced by extending the theory to a language genetically distinctive from English. Concentrating on only one node revealed that even a single phraseological unit containing the keyword could have many contextual features: collocational, colligational associations, and also rather abstract associations beyond the lexical level, namely semantic association.

The following sections will summarise the major findings pertinent to the three research questions set in Chapter 1 and insights gained from reflections on those results (Section 9.2). Discussions concerning the theoretical and methodological contributions made by this project will be provided in Section 9.3. Section 9.4 will discuss the pedagogical implications of the findings for English language teaching in China. Finally, Section 9.5 will discuss the limitations of the project and offer suggestions for further research in relevant fields.

9.2 Major findings

Before summarising the major findings of this project, we shall revisit the research questions formulated in Chapter 1:

1. Can collocation, colligation and semantic association as defined in the Lexical Priming Theory account for the range of the lexical and syntactical behaviour of Chinese?

2. If the answer to the first research question is positive, in what ways do English and Chinese differ or show similarity with regard to the characteristic primings for lexical and syntactical behaviour?

3. Based on any lexical and syntactical similarities and differences that we found between English and Chinese, what advances or alterations could we provide for the Lexical Priming Theory?

Taking the first two questions together, our ultimate goal was to test whether the basic concepts defined in LPT could also be applied to a Chinese node and how valid this theory was for this node.

The departure point of investigating 世界 *shi4jie4* was its collocation. It was not surprising to find collocation exists in Chinese, as this phenomenon has been proved by a number of cross-linguistic researchers (e.g., Xiao and McEnery 2006); what makes this present study significant is that it suggests how collocation is primed in Chinese. Taking Chinese morphological features into consideration, we have shown, using evidence detected from 世界 *shi4jie4,* that Chinese can be primed both in terms of single character and as a nesting. In addition, we propose that nesting can be applied as a basic corresponding pattern and departure point for cross-linguistic study

between English and Chinese.

Another theoretical contribution made by this project is that it has provided evidence of the existence of colligation in Chinese by undertaking a comprehensive examination for the single node 世界 *shi4jie4* and word sequences containing the node at the group level, grammatical function level, and textual level (positions in a sentence). Both standard colligational and co-colligational analysis were applied in the project. The former was applied to descriptive statistics (essentially probabilities) in the analysis of some aspects of the colligational behaviour of 世界 *shi4jie4* (and also *world*) and frequently occurring nestings containing the nodes. Co-colligational analysis was used to describe the relationship of attraction or repulsion between different syntactic slots associated with the lexical behaviour of the node or particular nesting containing the node. Simply put, our co-colligational analysis was focused on the examination of collocational behaviour within different colligational patterns. This method moved from intra- into inter-colligation by measuring the probability of each collocation occurring within each grammatical position under a particular grammatical function. Results presented in Chapters 6 and 7 point towards a relationship between the collocation and colligation of a node. The behaviour of 世界 *shi4jie4* (and also *world*) was biased towards particular syntactic slots (grammatical positions) and against others when serving a particular grammatical function. The evidence also suggests that the preferences/avoidances are not always characteristics of the node's syntagmatic behaviour and may interact with other such factors as collocation and semantic association. The complexity of the patterns identified in the study point to an interaction between units at different levels (e.g., collocation, colligation and semantic association).

In traditional modern Chinese grammar books (e.g., *Ma's Grammar Book*), modern Chinese is described as a parataxis language whose loose syntactical formality is linked with semantic meanings. However, the interplay and dependency relationship found in this study shows that the "parataxis language" label applied to modern Chinese might be too extreme. The semantically connected language also demonstrates syntactical preference/aversion either within a lower group, such as the position within a nominal group, or within a higher rank, such as the grammatical function a particular node serves. An examination of collocation at a greater distance (e.g., within 15 tokens) showed that collocation of 世界 *shi4jie4* can also operate with more distant neighbours. Distant collocates can be grouped into several semantic associations. For example, the collocates of the nesting 震惊世界 *zhen4jing1* (to shock the world) demonstrated five sub-semantic associations to CHANGE (see Chapter 7), two sub-semantic associations to DISCOVERY, and two sub-semantic associations to DISASTER. Extending the examination to 15 tokens revealed that textual semantic associations existed; for example, a textual semantic association of CRITICISM existed at a greater distance when the close collocates were associated with the meaning of NEW POLICY (under the semantic association of CHANGE). These findings, though needing more evidence, imply that a grammatical system and frame for the description of Chinese grammar as a whole may need to be drawn concerning the interrelations of the categories.

In answering the second question, we compared the characteristic of the lexical and syntactical features found for *world* and 世界 *shi4jie4*. We first compared their lexical features by examining collocates on both sides within a word span of ±4. Collocates at positions L1 to L4 and R1 to R4 were compared respectively. Although similar

collocates, such as today's/当今 *dang1jin1* were found to occur at the same position (L1) for *world* and 世界 *shi4jie4*, most collocates were distinctive for the two nodes at the same positions. However, we found a number of collocates (and sometimes also colligates) were shared by *world* and 世界 *shi4jie4* even if they did not occur at the same position. For example, the prepositions *in* and *around* occurred at the L2 position to *world*, while their Chinese equivalent (在 zai4) occurred at the L1 position to 世界 *shi4jie4*.

Similar situations occurred when we compared the general colligational profiles of *world* and 世界 *shi4jie4*. No significant similar preference or avoidance for a particular grammatical pattern was found, except for a positive colligation between the nodes and PRESENT tense. Since we undertook co-colligational analyses of both *world* and 世界 *shi4jie4*, it was also necessary to compared their lexical and syntactical behaviour within complex situations. Therefore, we next conducted a holistic comparison across co-textual levels. In examining collocational behaviour within colligational patterns, we found the congruence in lexical behaviour lay in the divergence of grammatical patterns. Traditionally, cross-linguistic studies between English and Chinese have focused on comparisons within a single unit, such as collocation. Although more conceptual and abstract associations than collocations, such as colligations (e.g., Shao 2017), semantic preference and prosody (e.g., Wei and Li 2014, Xiao and McEnery 2006) have been examined, few analyses have been conducted at multi-textual levels. Accordingly, we suggest a more comprehensive comparison be undertaken using data found in this project, as the characteristic features of English and Chinese were reflected at all levels of phraseology rather than

just with one unit, such as collocation.

Our third goal was to advance and extend the use of priming. The next section aims to answer the last research questions explicitly, based on the lexical and syntactical similarities and differences we found between English and Chinese, as the advances or extensions made in this project offer both theoretical and methodological implications for further research, especially for the study of Chinese.

9.3 Theoretical and methodological implications

Since first published in 2005, Lexical Priming Theory has seemed to be a revelation, because the ideas it presented provided answers to a number of linguistic issues. Pace-Sigge (2017: xi) described the theory as "an alternative lexically-based and natural-language usage driven approach to language research." Concerned, in part, with how naturalness is achieved and how an explanation of nature would impinge on what is possible, LPT provides insights on the "driving force" behind a language. The theory has progressed by being taken up by an increasing number of linguists working in various fields, who have offered modifications, advances, and varieties of its application. This project, based on previous demonstrations of the theory's influence, showcases some of applications of and advances to LPT by looking at the patterns and behaviours of a Chinese node. Theoretically, this project shows the potential of using LPT to describe Chinese linguistic features, especially for grammar. Compared with English, whose grammar system was established very early on, Chinese grammar seems rather loose. Modern Chinese grammar was formulated in reaction to the introduction of English grammar systems, and basic

Chinese grammatical categories then established in accordance with ideas of English grammar. Indeed, modern Chinese grammar is much more complex than is generally thought. The thought-pivot feature enables Chinese to put forward a topic at the beginning, followed by comments linked with tight semantic meaning but rather loose grammatical structures. Every kind of word structure can be found in Mandarin Chinese, which complexity yields an unclear demarcation of "words." Traditional descriptions of modern Chinese grammar do not clearly delineate the typical morphological unit of Chinese (character) from the designated unit from English (word). Both single characters and combinations of two or more characters can be entered in a Chinese dictionary. Using the data examined in the present study, we showed the inconvenience and unfeasibility of using the English term "word" to define various combinations of Chinese characters. At the same time, we demonstrated the potential validity of using *nesting* as convenient shorthand to describe complex Chinese word sequences. In addition, we proposed that LPT might serve as a theoretical backbone for the construction of modern Chinese grammar. Hoey argued (2005:169) that "The grammatical category we assign to a word…is simply a convenient label we give to the combination of (some of) the word's most characteristic and genre-independent primings." Thus, the statement that *world* typically operates as a noun is really an accumulation of its collocations, colligations, and semantic associations. Having shown the phenomenon of the above three fundamental concepts in Chinese, we can, in the next stage, describe Chinese grammar as an outcome of the interaction of the three factors. The category a Chinese character or nesting belongs to is its grammatical priming, which is a matter of tendency rather than a requirement.

Another theoretical and methodological effort we made was to explore an aspect of colligation not yet tackled for Chinese, co-colligational analysis; i.e., the description of a word's or word sequence's colligational behaviour at more than one level. Examining colligations at a co-textual level has not even been considered for English, even though a combinatory grammatical profile of *consequence* has been provided by Hoey (2005). Our qualitative analysis showed that the specific property of a given word or a word sequence depends on two or more levels; for example, an interaction between collocations and colligations. The existence of such interplay between collocation, colligation and semantic association poses a new methodological challenge to corpus lexis study, as this level of complexity cannot be dealt with by traditional colligational description. More levels need to be considered together when examining the colligational features of a word or word sequence.

9.4 Pedagogical implications

Teaching and learning a second language are necessarily different from teaching and learning a first language. The transfer of primings from earlier to later language is unavoidable, with first language primings being subconsciously activated to impose primings in second language learning.

Observing Chinese grammar from a cross-linguistic perspective has implications for learning English as a second language in China. Their shared semantic association shows that both English and Mandarin speakers recognise the *world*, to a degree, in a similar way. For example, they all prefer to apply *world* to represent the highest rank of evaluation, even if Chinese speakers show a stronger preference for and employ

more varies word choices to express this meaning. However, their different primings for colligational use or collocate choice may explain why sometimes L2 learners produce a "non-idiomatic" language, such as "learn the knowledge," as exemplified in the first chapter. In the process of priming this seemingly non-native-like collocation, the learners subconsciously translated their first language priming to their second language. As discussed in Chapter 1, 学习知识 *xue2xi2zhi1shi* (to obtain the knowledge) is a fairly frequent collocation in Mandarin, where 学习 *xue2xi2* can be translated directly to *learn* in English (per the bilingual dictionary) and 知识 *zhi1shi* can be translated literally into *knowledge*. After translating the word sequence from L1 to L2, a self-reflexive grammar was applied to check the correctness of this usage. The V+N grammar confirmed the students' translation, and the learners were satisfied with the product. Shared semantic association and colligation were selected in this process, but improper collocates were chosen to fill in the grammatical opportunity.

A word may have a number of collocates and semantic associations that may, in turn, vary when appearing in different grammatical functions. The transfer of primings from L1 to L2 is unavoidable and word-for-word translation is a helpful early-stage strategy for language learning. However, as a learner's second language proficiency improves, the translation strategy is suggested to give way to encountering new words in different contexts or situations, enabling the leaner to either reinforce or crack his/her primings to perfect his/her language learning. Despite the obvious differences arising from second language learning, this does not mean second language learners shall do exactly as native-speakers do. Indeed, there is no set of agreed-upon primings that either a native or non-native speaker should acquire.

Priming is, after all, unique to the individual. What we claim here is that neither simple word-for-word translation nor rote recitation of a long word list are wise strategies for second language learning, as a list or word entry in a dictionary may strip words of their primings. For language learning, it is suggested that priming be recalled from the context of a language learner encountering evidence and generalising from authentic materials.

For some English teachers in China, English learning is likely to be a matter of memorizing long lists of words. However, vocabulary wordlists are likely to introduce isolated meanings of individual words by showing one or two word-to-word translations. Lexical syllabi and wordlists focus on the most common and most frequently occurring words, while ignoring their applications in particular contexts; i.e., their collocational, colligational and semantic associational primings in L2. Reciting words in isolation from their priming features will only improve the breadth of learners' vocabulary knowledge, not its depth. The practice of learning words by lists rather than contexts will definitely worsen the situation, as a "word list strips the primings of the words and asserts the primings of its strict parallelism" (Hoey, 2005:184)

The comparison of *world* and 世界 *shi4jie4* in this project showed that the priming features of English and Chinese may exist at a co-textual level. Examining only one pair of nodes revealed a rather complex co-textual feature that may require language learners to possess several skills. Therefore, an explicit and holistic teaching approach is called for. For both learners and teachers of English, deviations in collocation and semantic association may cause non-idiomatic production, even if

word combinations are grammatically correct.

The teaching strategy we suggest for second language teaching at the preliminary stage is to provide students with single focused and generalising collocations, colligations and semantic associations of new words. This can be done in a multitude of ways, such as drilling exercises, texts, or tapes with repeated instances of a word sequence. For most L2 learners, the classroom and the teaching materials selected by their teacher provide the only context for their L2 priming, so it is essential that these primings are helpful in the area in which they will be used. This requires teachers and schools to choose the most appropriate textbooks and reading materials to help students form correct L2 primings. Primings are the results of a speaker encountering evidence and generalising from it. The most efficient and useful shortcuts to primings, especially at the early stage of L2 learning, are repetition and recitation. However, the materials repeated or recited must be authentic and have certain target language features.

The degree of difficulty of learning a target language is believed to depend preliminarily on the extent to which the L2 language pattern is similar to or different from that of the L1 language. Therefore, it is better for teachers to raise students' awareness of the occasional alien usages of semantic associations and colligations. With regard to usage differences across the two languages, both learners and teachers are urged to be aware of collocational, colligational and semantic associational divergence. Corresponding equivalence resides in a more complex pattern than simple word-to-word translation. In the classroom context, teachers can act as a scaffolders by pointing out which primings are not in harmony with those of their

likely listeners or readers, and will therefore sound non-idiomatic to them. Teachers can crack inharmonious primings directly or recommend that students read more authentic materials to find the conflicts between the learners' own primings and the common primings shared by the community in which their listeners or readers live. This was the strategy applied by the Prince of Grammar introduced in the Chapter One, even if he did not know the theoretical basis of it.

It is also suggested that those learning outside of a classroom context should take an advantage of a corpus. Online free corpora of English and Chinese are sources of authentic data collected by linguists for various research purposes. Learners can access these corpora and retrieve hundreds of concordance lines to identify syntax patterns commonly associated with a keyword. Being exposed to language in use, the learners could either sum up the target language's characteristic priming features or noticing gaps between their L1 and L2 at a co-textual level. Learners are strongly urged to make use of concordancers designed specifically for finding L2 priming features and comparing the range of lexical and syntactical behaviours between L1 and L2—for example, the Prime Machine, reviewed in Chapter 2. Distinct from other concordancers, the Prime Machine highlights: the links between words and related words, thus helping language learners distinguish near-synonyms; alternative English translations of Chinese characters or nestings, which helps learners select appropriate words based on the semantic meaning they want to express; and the tendency for words or phrases to occur at different positions in a text. This feature helps learners go beyond the local context and find about textual semantic associations and colligations.

9.5 Limitations of the project and suggestions for further research

This thesis also has some obvious limitations, due to time and space constraints. The first concerns methodological rigour. Only one pair of words was investigated in the present study. More evidence needs to be found to determine the extent to which lexical priming is valid for and applicable to Chinese. Secondly, we only investigated data represented in corpora that indicate the kinds of data for which members of a particular speech community might be primed. However, a hypothesis concerning psychology ought to be examined using psychological machines or experiments. As Hoey emphasised[27] only through the study of human neurons and brains can the claims proposed in LPT be proved; otherwise, priming is only a hypothesis deduced from corpus data. Thus, future studies should include neuron tests or psychological experiments designed to evidence the theory.

Priming is unique to the individual, and there is no agreed-upon set of primings that a learner should acquire. What distinguishes learners is not therefore whether they were native or non-native, "but how [their] primings come into existence" (Hoey 2005:184). Therefore, it is suggested the study of LPT be extended to the learners' language's corpus, to find characteristic priming features shared by language learners and determine the extent to which learners' L1 primings influence their L2 use.

This present project represents only a first, preliminary step towards identifying a fuller picture of the validity and applicability of lexical priming for Chinese. Other concepts proposed in the theory need to be examined for Chinese in further studies.

---

[27] In a Skype meeting with the candidate.

References

Altenberg, B. (1999). 'Adverbial connectors in English and Swedish: Semantic and lexical correspondences.' In H. Hasselgård and S. Oksefjell (eds.), *Out of Corpora. Studies in Honour of Stig Johansson* (pp.249-268). Amsterdam&Atlanta: Rodopi,

Baker, H., McEnery, T. and Hardie, A. (2017). 'A corpus-based investigation into English representations of Turks and Ottomans in the early modern period.' In M. Pace-Sigge and K.J. Patternson (ed.), *Lexical Priming: Applications and Advances* (pp.41-66). Amsterdam: John Benjamins Publishing Company.

Bastow, T. (2003). Friends and allies: binomials in a corpus of US defence speeches, paper given at ASLA Symposium, Orebro University, 6 Novermber.

Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistic and Linguistic Theory4-2:119-139.*

Benson, M. (1985). Collocations and Idioms. In R. Ilson (Ed.). *Dictionaries, Lexicography and Language Learning.* Oxford: Pergamon.

Bolinger, D. (1980). *Language – the loaden weapon, the use & abuse of language today*. London: Longman.

Bublitz, W. (1995). Semantic prosody and cohesive company: Somewhat predictable. *General and Theoretical Papers 347. Duisburg: L.A.U.D. (Linguistic Agency University of Duisburg): 1-23.*

Chen, H.W. (1998). 汉英翻译基础 *Chinese-English Translation*. Shanghai: Shanghai Education Press.

Chao, Y.R. (1945). The logical structure of Chinese words. *Language: xxii 1.*

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1958). Paper given at *Third Texas conference on problems of linguistic analysis in English*. Austin: University of Texas.

Deterding, D. (2006) The pronunciation of English by speakers from China. *English World-Wide (27):175-198.*

Doyle, P.G. (2003). Replicating corpus linguistics: a corpus-driven investigation of

lexical networks in text, unpublished Ph.D thesis, University of Lancaster.

Ellis, N.C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics (32): 17-44.*

Ellis, N., Römer, U. and O'Donnell, M.B. (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigation of Construction Grammar.* Michigan: The University of Michigan Press.

Fillmore, C. (1992). 'Corpus linguistics' or 'Computer-aided armchair linguistics'. In J. Svartvik (eds.) *Directions in Corpus Linguistics* (pp. 35-60), Proceedings of Nobel Symposium 82, Stockholm 4-8 August 1991. Berlin: Mouton.

Firth, J.R. (1957). *Papers in Linguistics, 1934-1951.* London: Oxford University Press.

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

Francis, W. N. (1982). 'Problems of assembling and computerising large corpora'. In S. Johansson (eds.) Comp*uter Corpora in English Language Research* (pp.7-24). Bergen: Norwegian Computing Centre for the Humanities.

Francis, W. N. (1985). After-dinner speech at the 5th ICAME conference at Windermere. *ICAME news,* 10 (5).

Frei, H. (1941). Un systeme chinois des aspects. *Acta Linguistica, II, 137-150.*

Fung, L. and R. Carter. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings.'*Applied Linguistics 28(3)*: p. 410-439.

Ge, X.H. (2014). 翻译与语言发展：'词汇触发'理论视角及其分析 Translation and language change: Analysis from the perspective of 'Lexical Priming' theory. *Foreign Languages and Their Teaching* (4): 15-20.

Goldberg, A.E. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Greenbaum, S. (1974). Some verb-intensifier collocations in American and British English. *American Speech (49): 79-89.*

Greaves, C. (2007). 'Software demonstration'. *Keyness in Text*, Certosa di Pontignano, University of Siena, Italy, 26-30 June.

Gross, M. (1993). 'Local grammars and their representation by finite automata'. In M. Hoey (ed.) *Data, Description, Discourse* (pp.26-38). London: Harper Collins.

Harley, B. (1986). *Age in Second Language Acquisition.* Clevedon: Multilingual Matters.

Halliday, M.A.K. (1959). *The language of the Chinese 'Secret History of the Mongols'*. Oxford: Blackwell.

Halliday, M.A.K. (1976). *System and Function in Language.* Oxford: Oxford University Press.

Halliday, M.A.K. and R. Hasan. (1976). *Cohesion in English*. London: Longman.

Halliday, M.A.K. & James, Z.L. (1993) 'A quantitative study of polarity and primary tense in the English finite clause' In J. Sinclair, M. Hoey & G. Fox (ed.). *Techniques of Description: Spoken and Written Discourse. A Festschrift for Malcolm Coulthard* (pp. 32-66). London: Routledge.

Halliday, M.A.K. (2007). 'Medieval Chinese grammar'. In Jonathan J. Webster (eds.) *Studies in Chinese Language* (pp.3-162). Beijing: Peking University press.

Hanks, P. (2013). *Lexical Analysis*. London: John Benjamins.

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

Hoey, M. (1991). *Patterns of Lexis in Text.* Oxford: Oxford University Press.

Hoey, M. (1997a). Lexical problems for the language learner (and the hint of a textual solution), in *Proceedings of the 5th Latin American ESP Colloquium*, Merida, Venezuela.

Hoey, M. (1997b). From concordance to text structure: New uses for computer corpora. *In PALC'97: Proceedings of practical applications of linguistic corpora conference (p2-23).* University of Lodz

Hoey, M. (2003). Lexical priming and the properties of text. URL: www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm. accessed on August 23, 2014.

Hoey, M. (2005). *Lexical Priming: A new theory of words and language.* London: Routledge.

Hoey, M. (2011). 'Lexical priming and translation'. In A. Kruger, K. Wallmarch & J. Munday (ed.). *Corpus-based Translation Studies: Research and Applications* (pp.153-168). London: Continuum International Publishing House.

Hoey, M. and J. Shao. (2015). 'Lexical priming: The odd case of a psycholinguistic theory that generates corpus-linguistic hypotheses for both English and Chinese' in B. Zou, S. Smith and M. Hoey (ed.). *Corpus Linguistics in Chinese Context* (pp.15-34). Hampshire: Palgrave Macmillan

Hundt, M., A. Sand, and R. Siemund. (1998). *Manual of Information to Accompany the Freiburg-LOB Corpus of British English.* Freiburg: University of Freiburg

Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language 2: 133-158.*

Hunston, S., Francis, G. & Manning, E. (1997). Grammar and vocabulary: Showing the connection. *ELT Journal, 51(3): 208-216.*

Hunston, S. and Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied linguistics 19(1): 45-72.*

Hunston, S. & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English.* Philadelphia, PA: Benjamins.

Hunston, S. (2001). Colligation, lexis, pattern and text. In Scott and Thompson (ed.). *Patterns in Text: In Honour of Michael Hoey* (pp.13-33). Amsterdam: John Benjamins Publishing Co.

Hunston,S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics* 12(2): 249-268.

Hornby, A.S. (1954). *A Guide to Patterns and Usage in English.* London: OUP.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics 19(1):24-44.*

Jaeco, S. (2017). 'Concordancing lexical primings: The rationale and design of a

user-friendly corpus tool for English language teaching and self-tutoring based on the Lexical Priming theory of language (pp.273-296).' In M. Pace-Sigge and K.J.Patternson (ed.), *Lexical Priming: Applications and Advances*. Amsterdam: John Benjamin Publishing Company

Jantunen, J. H. (2017). 'Lexical and morphological primings: A holistic phraseologicalanalysis of the Finnish time expression *kello'* In M. Pace-Sigge and K.J. Patternson (ed.), *Lexical Priming: Applications and Advances*(pp. 253-272). Amsterdam: John Benjamin Publishing Company

Johansson, S. and Hofland, K. (1994). 'Towards an English-Norwegian Parallel Corpus.' In *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Langue Research on Computerised Corpora, Zurich 1993*, U. Prles, G. Tottle and P. Schneider (ed.). pp.25-37. Amsterdam: Rodopt.

Jones, S., & Sinclair, J. (1974). English lexical collocations: a study in computational linguistics. *Cahiers de lexicologie 24: 15-61.*

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Kilgarriff, Adam, Rychlý，Pavel, Smrž，Pavel & Tugwell, David (2004) The Sketch Engine Proceedings of EURALEX.

http://www.euralex.org/elx_proceedings/Euralex2004/ p105-114

Kilgarriff, A., N. Keng, & S. Smith (2015) 'Learning Chinese with the Sketch Engine' In B. Zou, S. Smith, & M. Hoey (ed.). *Corpus Linguistics in Chinese Contexts,* p63-73. Hampshire: Palgrave macmillan

Kjellmer, G. (1984). 'Some thoughts on collocational distinctiveness.' In J. Aarts and W. Meijs (ed.) *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi.

Kjellmer, G. (1987). 'Aspects of English collocations.' In W. Meijs (eds.) *Corpus Linguistics and Beyond. Proceedings of the Seventh International Conference on English Language Research on Computerised Corpora* (pp.133-140). Amsterdam: Rodopi.

Langendoen, D.T. (1968). *The London School of Linguistics: A study of the Linguistic*

Theories of B. Malinowski and J.R. Firth. Cambridge, MA: MIT Press.

Leech, G.N. (1966). *A Linguistic Guide to English Poetry*. London: Longman.

Leech, G.N. (1974). *Semantics*. Harmondsworth: Penguin.

Leech, G.N. (1991). 'The state of the art in corpus linguistics.' In K. Aijmer and B. Altenberg (ed.) *English Corpus Linguistics* (pp.8-29). London: Longman

Leech, G.N. (1992). 'Corpora and theories of linguistics performance.' In J. Svartvik (eds.) *Directions in Corpus Linguistics* (pp. 105-122). Proceedings of Nobel Symposium 82, Stockholm 4-8 August 1991. Berlin: Mouton.

Leech, A. (2001). 'The role of frequency in ELT: new corpus evidence brings a re-appraisal.' *Foreign Language Teaching and Research 33(5): 328-339.*

Lewandowska-Tomaszczyk, B. (1996). 'Cross-linguistic and language-specific aspects of semantic prosody.' *Language Science, (18):153-178.*

Li, C.N. and S.A. Thompson (1989). *Mandarin Chinese: A Functional Reference Grammar*. Los Angeles: University of California Press

Li. L. (2006). 《词汇触发：有关词汇和语言的新理论》评介 A review on Lexical Priming: A New Theory of Words and Language. *Foreign Language Teaching and Research. 38(3): 231-233.*

Li, W.Z. and S. Smith. (2015). "Introduction" In B. Zou, S. Smith and M. Hoey (ed). *Corpus Linguistics in Chinese Context* (pp.1-14). Hampshire: Palgrave Macmillan.

Louw, Bill. (1993). 'Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies.' In Mona Baker, Gill Francis and Elena Tognini-Bonelli (ed.), Text and Technology: In Honour of John Sinclair (p157-176). Amsterdam: John Benjamins.

Louw, Bill. (2000). 'Contextual prosodic theory: Bringing semantic prosodies to life.' In Chris Heffer and Hellen Saunston (ed.), *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham.

Meyer, C.F. (2002). *English Corpus Linguistics*. Cambridge: Cambridge University Press.

McEnery. A. & Wilson, A. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh

University Press.

McEnery, T., Xiao, R. and Mo, L. (2003). Aspect marking in English and Chinese: using the Lancaster Corpus of Mandarin Chinese for contrastive language study. *Literary and Linguistic Computing 18(4): 361-378.*

McEnery, A. & Z. Xiao (2007) Quantifying constructions in English and Chinese: A corpus-based contrastive study. Paper presented at Corpus Linguistics 2007. p.28-30 July 2007, Birmingham University.

Mullie, J. (1970). *The Structural Principles of the Chinese Language*, Vols I-II, (Peiping, Peit'ang Lazarist Press).

Neely, J.H. (1976). Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory process. *Memory & Cognition 4(5)*: *648-654.*

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology. General 106: 226-254.*

O'Hallorab, K.A. (2007). Critical discourse analysis and the corpus-informed interpretation of metaphor at the register level. *Applied Linguistic 28:1-24.*

Pace-Sigge, M. (2013). *Lexical Priming in Spoken English Usage*. Hampshire: Palgrave Macmillan.

Pace-Sigge, M. (2017).Can lexical priming be detected in conversation turn-taking strategies In M. Pace-Sigge and K.J. Patternson (ed.), *Lexical Priming: Applications and Advances* (p93-120). Amsterdam: John Benjamin Publishing Company.

Palmer, F.R. (1975). *The English Verb*. London: Longman.

Palmer, H.E. (1933). *Second Interim Report on English Collocations*. Tokyo: Institute for Research in English Teaching.

Partingon, A. (1998). *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Partingon, A. (2004). 'Utterly content in each other's company: Semantic prosody and semantic preference'. *International Journal of Corpus Linguistics 9(1):131-156.*

Patterson, K.J. (2016). 'The analysis of metaphor: To what extent can the theory of

lexical priming help our understanding of metaphor usage and comprehension?' *Journal of Psycholinguistic Research 45(2)*: 237-258.

Patterson, K.J. (2017). 'Lexical priming and metaphor - Evidence of nesting in metaphoric lanague.' In M. Pace-Sigge and K.J. Patternson (ed.), *Lexical Priming: Applications and Advances* (pp.141-162). Amsterdam: Benjamin Publishing Company

Quirk, R. and Svartvik, J. (1966). *Investigating Linguistic Acceptability*, *Janualinguarum, Series Minor 54.* The Hague: Mouton.

Renouf, A. & Sinclair, J. (1991). *Collocational Frameworks in English.* London: Longman

Ringbom, H. (1998). 'High frequency verbs in the ICLE corpus' In A.Renouf (eds.). *Explorations in Corpus Linguistics* (pp.191-200). Amsterdam: Rodopi.

Sardinha, T.B. (2000) Semantic prosodies in English and Portuguese: A contrastive study *Cuadernos De Filolog ú Inglesa, 9(1): 93-110.*

Scarborough, Don L., Cortese, Charles and Scarborough, Hollis, S. (1977). 'Frequency and repetition effects in lexical memory'. *Journal of Experimental Psychology: Human Perception and Performance 3(1):117-134.*

Shao, J. (2017). 'Teaching near-synonyms more effectively: A case study of "happy" words in Mandarin Chinese. 'In M. Pace-Sigge and K.J. Patternson (ed.), *Lexical Priming: Applications and Advances* (pp.163-188). Amsterdam: John Benjamin Publishing Company.

Sinclair, J. (1996). 'Beginning the study of lexis.' In C.E. Bazell, J.C. Catford, M. A.

K. Halliday and R.H. Robins (ed.) *In Memory of J.R. Firth* (pp.410-430). London: Longman.

Sinclair, J. (1987). *Looking up: An account of the COLBUILD project in lexical computing and the development of Collins COLBUILD English language dictionary.* London/Glasgow: Collins.

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

Sinclair, J. (1992). 'Trust the text'. In M. Davies and L. Ravellie (ed.): *Advances in*

Systemic Linguistics (pp. 5-19). London: Printer.

Sinclair, J. (1996). "The search for units of meaning." *Textux (9): 75-106.*

Sinclair, J. and Renouf, A. (1998). 'A lexical syllabus for language learning.' In R. Carter and M. McCarthy (ed.) *Vocabulary and Language Teaching* (pp.140-160). London: Longman.

Sinclair, J. (2004a). *Trust the Text.* London: Routledge.

Sinclair, J. (2004b). 'Progress and prospects in corpus linguistics.' *Modern Foreign Languages 27(2): 112-128.*

Sinclair, J. (2005). The phrase, the whole phrase and nothing but the phrase. Plenary lecture, *Phraseology*, 2005, Louvain-la-Neuve, October.

Stubbs, M. (1995). Collocations and semantic profiles: on the cause of trouble with quantitative methods. *Function of Language:2(1):1-33.*

Stubbs, M.(1996). *Text and Corpus Analysis.* Oxford: Blackwell.

Stubbs, M. (2001). *Words and Phrases: Corpus studies of lexical semantics.* New York: Blackwell.

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics 7(2):215-244.*

Stubbs, M. (2008).Quantitative data on multi-word sequences in English: the case of word *world* In M. Hoey, M. Mahlberg, M. Stubbs and W. Teubert (eds). *Text, Discourse and Corpora* (pp.163-190). London: Continuum.

Stubbs, M. (2009). Memorial Article: John Sinclair (1933-2007). The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics 30(1): 115-137.*

Su, N.N. (2012). 词汇触发理论及对大学英语词汇教学的启示 Lexical Priming and Its Implication on College English Vocabulary Teaching. *Journal of Sichuan College of Education. (28) 28: 97-101.*

Svartvik, J. (1996). Corpora are becoming mainstream. In J. Thomas and M. Short (ed.) *Using Corpora for Language Research* (pp.3-13). London: Longman.

Svartvik, J. (2007) Corpus linguistics 25+ years on. In R. Facchinetti (eds) *Corpus Linguistics 25 Years on* (pp.11-26). Amsterdam: Rodopi.

Tao H.Y. (2003) Toward an emergent view of lexical semantics. *Language and*

Linguistics, 4(4): 837:856.

Taylor, J.R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

Thomas, J., & Short, M. (1996). *Using Corpora for Language Research.* London: Longman.

Tyler, A. (2010). 'Usage-based approaches to language and their applications to second language learning' In A. Mackey (eds) *Annual Review of Applied Linguistics* (pp 270-291). Cambridge: Cambridge University Press.

Warren, M. (2007). Making sense of phraseological variation. Plenary lecture, Keyness in Text, Certosa di Pontignano, University of Siena, Italy, 26-30 June.

Wei Naixing. (1999). Towards defining collocations: A practical scheme for study of collocations in WAP texts. Unpublished doctoral dissertation, Shanghai: Shanghai Jiaotong University, China

Wei, N.X and X.H. Li. (2014). Exploring semantic preference and semantic prosody across English and Chinese: Their roles for cross-linguistic equivalence. *Corpus Linguistics and Ling 10(1):103-138.*

Wei, N.X. (2015). 'Foreword'. In B. Zou, S. Smith and M. Hoey (ed.). *Corpus Linguistics in Chinese Context* (pp.xiii-xvii). Hampshire: Palgrave Macmillan

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistic 10(3): 283-305.*

Xiao, R. and T. McEnery. (2006). Collocation, semantic prosody and near synonym: A cross-linguistic perspective. *Applied Linguistic 27(1):103-129.*

Xiao, R. (2007). What can SLA learn from contrastive corpus linguistics? The case of passive constructions in Chinese Learner English. *Indonesia Journal of English Language Teaching 3(1): 1-19.*

Xiao, R. (2011). Word clusters and reformulation markers in Chinese and English. *Language in Contrast 11(2): 145-171.*

Xiao, R. and G.R. Dai. (2014). Lexical and grammatical properties of translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics and Ling 10(1): 11-55.*

Yang, Huizhong. (2002). *An Introduction to Corpus Linguistics. Shanghai:* Shanghai

Foreign Language Education Press.

Zhang,W. (2009). Semantic prosody and ESL/EFL vocabulary pedagogy. *TESL Canada Journal:26(2):1-12.*