A Corpus-Based Study on the Use of Phrasal Verbs by L1 Users and Chinese EFL Learners

by

JIE LIANG

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics and Communication

School of English, Drama and Creative Studies

College of Arts and Law

University of Birmingham

October 2024

Abstract

Many Chinese EFL learners have faced challenges in learning and using English phrasal verbs (PVs). One reason learners struggle to use PVs naturally and frequently could be the difficulty of identifying them. The traditional approach to PV identification typically focuses on the grammatical function of the components, testing whether the particle is adverbial or prepositional (i.e. prepositional verbs). However, the traditional approach fails to show the semantic feature of PVs, especially how the meanings of a PV are established.

This study takes a semantic approach to identify and classify PVs in the Bank of English Corpus and examines verb + particle combinations that have non-compositional meanings. While some verb + particle combinations (e.g. *come across, look for*) may not be traditionally identified as 'phrasal', this study regards them as PVs because corpus evidence indicates that the meaning of these combinations cannot be predicted by their components. The study then investigates how L1 users use PVs and focuses particularly on lexico-grammatical patterns associated with specific PVs in the Bank of English Corpus. In addition, the study also uses a Chinese learner corpus and a specialised reference corpus to compare the use of PVs between L1 users and Chinese EFL learners. Results show that Chinese EFL learners underuse PVs in their writing. The overall frequency of PVs is lower in the Chinese learner corpus. Learners have difficulty using PVs, especially the non-contiguous form of PVs, and tend to make mistakes due to a lack of phraseological knowledge.

This study also discusses the difficulties of replacing PVs with one-word verbs from the perspective of pattern grammar. Although it is possible to find a potential one-word

alternative that has an associated meaning with a PV, corpus evidence shows that the grammatical patterns in which the PV and the one-word alternative appear are often different. Results suggest that Chinese EFL learners tend to prioritize semantic equivalence when they find the replacement, paying less attention to the grammatical differences between PVs and one-word alternatives.

Acknowledgements

I would like to express my sincere thanks to my lead supervisor, Dr. Crayton Walker, for his invaluable guidance and support. I am sincerely thankful for his patience and the countless hours he dedicated to reviewing drafts and providing insightful feedback. His thoughtful advice has helped me a lot throughout the four years.

I would also like to extend my deepest gratitude to my parents for everything they have done for me unconditionally. They have always believed in me and stood by me, even when I doubted myself. Their endless love and support have been my greatest motivation. Knowing that they are always there for me means the world to me.

I am truly thankful for my friends: Xuan Chen, Zhenyi Wei, and Fei Xie. Their constant encouragement and companionship have made this journey a little easier and much happier. No matter where they are, thank you for being by my side, both near and far.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Corpora and dictionaries

BNC	British National Corpus
BoE	Bank of English
COCA	Corpus of Contemporary American English
OED	Oxford English Dictionary
SECCL	The spoken sub-corpus of the SWECCL
SWECCL	Spoken and Written English Corpus of Chinese Learners
TECCL	Ten-thousand English Compositions of Chinese Learners
WECCL	The written sub-corpus of the SWECCL

Others

EFL	English as a foreign language
ESL	English as a second language
KWIC	Key word in context
L1	The first language or mother tongue
L2	The second language being learned or acquired
OWVs	One-word verbs
PVs	Phrasal verbs

List of Conventions

Asterisk *

i. Used to mark that the sentence or phrase fail to make sense, e.g. *We'd all spent many years *going* big changes *through*.

Bold text

i. Used to show grammatical patterns, e.g. the pattern **V *across* n**;
ii. Used to highlight verb + particle combinations that appear in a given example, e.g. "She ***took off*** her backpack and placed it on a chair."

"Double quotation marks"

i. Used to cite other scholars' work, e.g. "the occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991, p.170).

Italics

i. Used to refer to words or phrasal verbs under analysis or discussion, e.g. the meaning of *came across* in the example 'he *came across* a professional actor'.

'Single quotation marks'

i. Used to refer to the information in a given example, e.g. In lines 26 to 30, the objects that can be turned on are electronic devices, such as 'hair dryer', 'TV', 'radio', 'light', and 'computer'.
ii. Used to highlight terminology or specific labels, e.g. the term 'phrasal verbs' or the label 'free combinations'.
iii. Used to refer to the meaning of a word or a PV in the dictionary, e.g. *on* is sometimes used as an adjective, describing 'the state of an electronic device' (*OED online*).

Subscript

i. Used to denote different meanings of a word or a phrasal verb, e.g. *LOOK up*$_2$ refers to the second meaning of *LOOK up*.

Underlines

i. Used to highlight the contextual information that is under discussion, e.g. I ***admire*** him as an actor because he is completely generous in his work.

USE OF CAPITALISATION

i. Used to refer to lemmas (all word forms of a word). In this thesis, the lemma is used with the lexical verb in verb + particle combinations, e.g. *LOOK for* refers to all members of this combination, including *look for, looks for, looked for, looking for*.
ii. Used to refer to conceptual metaphors, e.g. GOOD is UP.

Chapter 1 Introduction

1.1 Background

1.1.1 Phraseology and corpora

Phraseology used to be a somewhat peripheral area in the study of the language (Sinclair, 2008; Granger, 2021). Before phraseology became more central in linguistic research, linguists focused primarily on grammar and lexis, regarding them as two separate but core systems in the language. Phraseology, however, integrates grammar and lexis and describes the language through a holistic way, which is distinctive from the tradition of language description. This is one of the reasons why the importance of phraseology has been consistently underestimated for so long (Sinclair, 2008). Another reason is that phraseology focuses on how words are combined to convey meanings (Sinclair, 2008). This is different from the traditional view that focuses on the grammatical rules of constructing a sentence. The phraseological view is well explained by the idiom principle (Sinclair, 1991), which highlights the restriction on the choice of words (see Section 3.1.3 for more details). The occurrence of one word is governed by the occurrence of other lexical items[1] in the surrounding context.

Conventionally, the study of phraseology was regarded as "a subfield of lexicography dealing with word combinations rather than single words" (Granger and Meunier, 2008, p.XIX). In this tradition, phraseological units[2] are classified into different types based

[1] Carter (2012) explains that lexical items can be single words or multi-word combinations, such as multi-word verbs (e.g. phrasal verbs) and idioms. The definition of lexical items in this thesis follows Carter's definition. In the current study, a lexical item refers to an individual word, such as *look*, or a multi-word combination conveying a specific meaning, such as *look forward to*.

[2] Cowie (1998) defines a phraseological unit as a group of words that are habitually used together to convey a particular meaning. This term specifically refers to multi-word combinations with a fixed or semi-fixed structure, such as idioms and restricted collocations.

on a number of linguistic criteria, including the non-compositionality of the meaning, the fixedness of the structure, the restriction on lexical choices, and the level of idiomaticity (Cowie, 1988; Howarth, 1996; Nesselhauf, 2004). This conventional approach to phraseological units has influenced the development of phraseology because it "[establishes] phraseology as a discipline in its own right" (Granger and Paquot, 2008, p.28). It also introduces new terminology associated with phraseology and provides researchers with a set of linguistic criteria to study and analyse phraseological units (Granger and Paquot, 2008).

Corpus-based research has accelerated the development of phraseology (Rundell, 2008). A corpus is a large collection of language data stored on a computer. It provides authentic examples of how language is used in both written and spoken discourse. Working with large amount of authentic data, linguists do not predefine linguistic criteria for phraseological units but identify them by examining the co-occurrence of lexical items (e.g. Sinclair, 1996). This is completely different from the conventional approach that relies heavily on linguists' intuitive judgment and experience (Granger, 2021).

In the corpus, if one word frequently co-occurs with another, they tend to form a lexical pattern. Drawing on the repeated lexical patterns found in the corpus, Sinclair (1991) argues that words do not occur together by chance; they co-select each other and constitute a single choice (see Section 3.1.3 for examples). Moreover, words are combined to form meaningful units not only at the level of collocation, but also at the

discourse level[3] (Sinclair, 1996; Stubbs, 2001). Compared with the conventional approach based on linguists' experience and intuition (Granger, 2021), this corpus-driven approach is able to identify a larger number of word combinations and has greatly expanded the range of phraseology (Granger and Paquot, 2008; McEnery and Hardie, 2012). Fixed, idiomatic word combinations, such as *kick the bucket*, used to be central to phraseology. However, a large amount of corpus evidence has shown that these highly restricted units are less frequent in the corpus (Moon, 1998). In addition, non-contiguous combinations and longer sequences of words can now be identified with the help of corpus (cf. Hunston and Francis, 2000).

Over the past decades, an increasing number of studies has enhanced our understanding of phraseology. Researchers has made considerable contributions by identifying different types of phraseological units (e.g. collocations, formulaic sequences, idioms, lexical bundles, n-grams, pattern grammar, etc.) and showing how these phraseological units convey meanings at both lexical and discourse levels (e.g. Cowie, 1988; Sinclair, 1996; Moon, 1998; Hunston and Francis, 2000; Wray, 2002; Biber, Conrad and Cortes, 2004). Phraseology has become more central in the field of linguistics and has gained a growing attention from language learners and teachers.

1.1.2 Teaching multi-word combinations in the ESL/EFL context

Multi-word combinations are prevalent in language (Nattinger and DeCarrico, 1992; Schmitt and Carter, 2004). Previous studies of phraseology have shown that experienced users of the language do not produce language from scratch; they have a

[3] The function or pragmatic meaning associated with the meaning unit can be identified at the discourse level.

large number of prefabricated word combinations available to them during use (Pawley and Syder, 1983; Sinclair, 1991; Wray, 2002). Pawley and Syder (1983) highlight the importance of multi-word combinations in their study and argue that being proficient in the language involves more than just understanding grammar rules and memorizing vocabulary. Schmitt (2000) agrees with this argument and suggests that language ability requires both the grammatical ability and the ability to use lexicalised word combinations.

In order to be more proficient in language use, non-native speakers need to develop two linguistic capabilities: nativelike selection and nativelike fluency (Pawley and Syder, 1983). The first is concerned with non-native speakers' ability to select expressions that are not only grammatically correct but also appropriate and frequent in the target context. The second refers to the ability to communicate in a natural and fluent manner. Pawley and Syder (1983) believe that native speakers' ability to produce natural and fluent language is closely associated with their "knowledge of 'sentence stem'[4] which are institutionalised or lexicalised" (p.191). Lexicalised word combinations connect vocabulary and grammar, helping learners understand how words are used together in real communication (Nattinger and DeCarrico, 1992).

Teaching language through a phraseological approach also helps learners understand the meanings of lexical items. Many corpus-based studies have shown that the meaning of one word is restricted by the surrounding words (e.g. Sinclair, 1996; Stubbs, 2001). Stubbs (2001) examined the collocation of *cause* and found that it frequently co-

[4] Sentence stems refer to "[units] of clause length or longer whose grammatical form and lexical content is wholly or largely fixed" (Pawley and Syder, 1983, p.191).

occurred with words carrying a negative meaning, such as *damage*, *problem*, etc. In other words, it is not possible to combine *cause* with *happiness* because the meaning of *cause* restricts the choice of surrounding words. Hunston and Francis (2000) suggested that the meanings of a polysemous word tended to occur in different grammatical patterns (see Section 3.2). Therefore, phraseology plays a crucial role in disambiguating polysemous items. Based on the surrounding lexical items and grammatical patterns, learners may find it easier to determine the intended sense of a word.

1.2 Rationale

The current study focuses on phrasal verbs (henceforth PVs). As a special but important member of multi-word combinations, PVs are ubiquitous in everyday communication. However, PVs have long been acknowledged as a significant challenge for ESL/EFL learners, especially for those whose first language (L1) does not have this special type of combinations (Dagut and Laufer, 1985; Liao and Fukuya, 2004; Waibel, 2007; Gilquin, 2015; Sung, 2020). The absence of PVs from Chinese poses a significant challenge for Chinese EFL learners (Lu and Sun, 2017; Wei, 2021). As a language learner, I also found learning PVs to be a challenging experience. Unlike one-word verbs, PVs often took multiple meanings which frequently led to confusion. These meanings tended to be hard to understand because they were not associated with the meanings of the individual components. As a result, the lack of PVs in learners' L1 and my personal experience of learning PVs motivated me to investigate how experienced users employ PVs in both written and oral communications and explore the major challenges that Chinese EFL learners encounter when they use these special multi-word combinations. Specifically, this study focuses on the phraseological use of PVs, especially collocation and grammatical patterns, and examines whether learning PVs

by focusing on collocation and grammatical patterns can help Chinese EFL learners understand the various meanings of PVs.

A PV is generally formed by a verb and a particle. Although it is hard to quantify the number of such verb + particle combinations in English, there is a certain trend that the majority of frequent PVs are constructed by high-frequency verbs, such as *get, take, make*, etc., and common particles, such as *up, down, out*, etc. (Gardner and Daivs, 2007). While the average learner will be exposed to these high-frequency verbs (Altenberg and Granger, 2001) and be familiar with the common particles, they still have difficulties using PVs because of the semantic and syntactic complexity of PVs (Dagut and Laufer, 1985; Liao and Fukuya, 2004). The semantic complexity can be reflected by the unpredictable meaning of PVs. Learners cannot predict the meaning of a PV based on the meanings of the verb and the particle (Hulstijn and Marchena, 1989; Side, 1990). In addition, the semantic complexity of PVs is also associated with the tendency that PVs usually have more than one meaning. A frequent PV in English may have 5.6 meanings on average (Gardner and Davis, 2007). Here are two examples showing the semantic complexity of a PV, *TAKE off*.

The plane ***took off***. (= to leave the ground and start flying)

He ***took off*** his jacket and sat next to her. (= to remove the clothes)

The two examples exemplify the polysemous nature of PVs, which makes it rather difficult for ESL/EFL learners (Side, 1990). In addition to the semantic complexity, the syntactic complexity of PVs lies in the unstable structure of PVs. If we look again at the two examples of *TAKE off*, the verb can be followed by the particle to form a contiguous combination (e.g. 'took off his jacket') or a non-contiguous combination (e.g. 'took his jacket off'). The movement of the particle does not affect the meaning

of the sentence. However, when the verb and the particle are separated by an intervening item, it would be harder for learners to recognise them as a meaningful unit (Gardner and Davis, 2018).

Given the semantic and syntactic complexity of PVs, it is important to find an effective way to teach PVs and alleviate ESL/EFL learners' difficulties of using PVs. A group of researchers focused on learners' behaviour of using PVs. Many ESL/EFL learners strategically avoid using PVs and they tend to choose one-word verbs which have an equivalent meaning (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Siyanova and Schmitt, 2007). There are various factors causing learners' avoidance of using PVs, such as the polysemous nature of PVs (especially PVs with idiomatic meanings), learners' proficiency level, the impact of L1, learning context, etc. (Dagut and Laufer, 1985; Liao and Fukuya, 2004; Schmitt and Redwood, 2011; Becker, 2014). Although researchers have concluded that learners tend to avoid using PVs and prefer one-word verbs, it remains unclear what approach can effectively enhance learners' understanding and use of PVs.

Another group of researchers explored possible ways of teaching PVs in the ESL/EFL context. However, it tends to be difficult to provide an effective method to teach PVs with various meanings (Schmitt, 2000). One of the widely used approaches is to create an inventory that includes the most frequent PVs based on the corpus data (e.g. Garnder and Davis, 2007; Liu, 2011; Garnier and Schmitt, 2015; Liu and Myers, 2020). To address the polysemous nature of PVs, these lists tended to give priority to the most frequent meaning(s) of PVs (e.g. Garnier and Schmitt, 2015; Liu and Myers, 2020). For example, Garnier and Schmitt (2015) established a list of 150 most frequent PVs and

provided the most frequent meaning of each PV. Language teachers and learners are advised to prioritise the most frequent meanings of PVs provided in the list, as high frequency is an indicator of significance (Garnier and Schmitt, 2015).

Based on the brief review of teaching applications, it is evident that researchers focus primarily on the semantic aspects of PVs, encouraging teachers and learners to pay attention to the most frequent meanings of PVs. While it is an effective approach to addressing polysemous PVs, it probably causes some problems. One of the problems is that learners become familiar with the frequent meanings but feel confused when they are encountering less frequent or peripheral meanings of PVs. Another potential weakness is associated with the lack of examples demonstrating the use of PVs in a context. The purpose of establishing lists of PVs is to enhance learners' lexical knowledge of PVs. However, Sinclair (1991) argues that lexis and grammar are not isolated systems of language. To encourage learners to use PVs more naturally and fluently, both lexical and grammatical knowledge of PVs should be considered. In addition to knowing the multiple meanings of a PV, learners also need to gain the phraseological knowledge of a PV and know how to use it properly. As a result, this study aims to examine the phraseological use of PVs based on the authentic corpus data, focusing particularly on the lexico-grammatical patterns associated with PVs. The study also intends to investigate Chinese EFL learners' behaviour of using PVs, identifying their difficulties of using PVs.

1.3 Aims and scope of the study

The primary aim of the current study is to identify the lexico-grammatical patterns associated with PVs and investigate how the use of PVs in Chinese EFL learners'

writing differs from the use in L1 users' writing. To achieve this aim, the study focuses on three objectives. The first objective is to investigate how experienced users of the language use PVs. To achieve this objective, the current research will focus on collocation, grammatical patterns and tense forms associated with specific PVs in the BoE corpus and therefore identify the phraseological use of them. The second objective is to investigate how Chinese EFL learners use frequent PVs in their writing and identify learners' difficulty with PVs. The investigation will compare the use of PVs in the Chinese learner corpus with the use in the reference corpus. The third objective is concerned with the strategies Chinese EFL learners use when they have difficulty using PVs. As discussed in Section 1.2, previous studies have concluded that EFL learners tend to avoid using PVs and replace PVs with one-word equivalents (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004). Therefore, the study will explore whether Chinese EFL learners also have such avoidance behaviour via an investigation into the use of specific PVs and their one-word equivalents in the Chinese learner corpus.

A secondary aim of the study is to develop semantic criteria that can be used to identify both two-word and three-word PVs. Traditionally, linguists and researchers often used syntactic tests to identify PVs (see Section 2.2.1). These tests were based on the syntactic features of PVs (see Section 2.1.3) and linguists used a top-down approach to differentiate between PVs and other multi-word verbs, especially prepositional verbs (e.g. Bolinger, 1971; Fraser, 1974; Biber et al., 1999). Although this top-down approach reflects the typical syntactic features of PVs, identifying them on the basis of these features is likely to ignore certain types of PVs, such as intransitive and inseparable ones (e.g. *come across*). In addition, these syntactic tests do not take three-word PVs

(e.g. *look forward to, put up with,* etc.) into account. Three-word PVs were often referred to as phrasal-prepositional verbs in many classroom materials or grammar coursebooks (e.g. Quirk et al., 1985; Biber et al., 1999; Downing and Locke, 2002). However, it was not clear why the second component was considered an adverbial particle, and the third component was a prepositional particle. The arbitrary labels on the particles may hinder learners from recognising and understanding three-word PVs (Celce-Murcia and Larsen-Freeman, 1999). Due to the weaknesses of syntactic tests, the current study will adopt a different approach, using semantic criteria to identify PVs and taking three-word PVs into consideration (see Chapter 5).

The scope of the study is restricted to frequent PVs. There are numerous PVs in English, but the higher frequency usually signals the importance of PVs (Gardner and Davis, 2007; Liu, 2011; Garnier and Schmitt, 2015). When investigating how L1 users use PVs, the current study pays attention to specific PVs containing a common and frequent verb *LOOK*. *LOOK* is one of the most frequent verbs in English. However, Altenberg & Granger (2001) noted that EFL learners have difficulty using those most frequent verbs proficiently. Another reason for choosing *LOOK* is that this verb is frequently combined with adverbial or prepositional particles to construct PVs (e.g. *LOOK up, LOOK down, LOOK forward to*, etc.). The examination of the lexico-grammatical patterning associated with these PVs is restricted to three aspects: (1) the nominal collocates of *LOOK*-PVs, (2) verb patterns associated with *LOOK*-PVs, and (3) the use of *-ing* form of *LOOK*-PVs. In terms of the comparison between Chinese EFL learners and L1 users, a learner corpus is created based on two available learner corpora of Chinese-speaking learners of English (see Section 4.2.2). The reference corpus used for comparison is based on three text files in the British National Corpus (BNC) (see Section 4.2.1). The

current study focuses on the use of 100 frequent PVs in the Chinese learner corpus and the reference corpus. Gardner and Davis (2007) identified the most frequent PVs in the BNC and produced a list of 100 frequent PVs (the full list of the 100 frequent PVs can be found in Appendix 1). This study is thus based on Gardner and Davis's list due to the frequency and significance of these 100 frequent PVs. In addition to the 100 frequent PVs, the study also identifies 20 frequent three-word PVs in the reference corpus (see the list of these 20 three-word PVs in Appendix 2) and compares the frequency of these three-word PVs in the two corpora. Finally, when analysing learners' choice between PVs and one-word verbs, it is not feasible to include all PVs and their one-word equivalents due to the large number of PVs in English. In order to limit the scope of the study, 10 PVs and their corresponding one-word equivalents are selected (see Section 4.3.3 for details). The study will compare the use of these ten pairs of verbs in the Chinese learner corpus and examine whether Chinese EFL learners are able to recognise the difference between PVs and one-word equivalents and use them properly in their writing.

1.4 Organisation of the thesis

This thesis is divided into nine chapters. I have outlined the background of the study in this chapter, explaining the importance of phraseology in language use and learners' difficulties with PVs. It also contextualises the study by explaining the rationale and defining the scope of the current research.

The subsequent two chapters establish the theoretical foundations for the current research. Chapter 2 aims to show the complexity of English PVs (e.g. the lack of a standard definition and the difficulty of identifying PVs). This chapter also discusses

the problems of teaching and learning PVs in the ESL/EFL context. Chapter 3 focuses on three aspects (i.e. collocation, pattern grammar, and tense form) that may affect the meaning of lexical items in English. Given the diversity of collocation and grammatical patterns, this chapter not only outlines the characteristics of the two phraseological aspects, but also explains how the current study deal with each of them.

Chapter 4 gives a detailed account of the methodology, including research questions, the corpora used to collect the data, and techniques used to analyse the data. Since the research is corpus-based, I used the Bank of English Corpus (BoE) to examine how L1 users use PVs with a particular focus on the lexico-grammatical patterns associated with PVs. Meanwhile, I also used a specialised sub-corpus of the BNC as the reference corpus for comparison. In terms of the learner corpus for comparison, the research employed a Chinese learner corpus which was based on two ready-made Chinese learner corpora. Chapter 5 can be seen as a part of the methodology, because this chapter introduces how the current study classifies verb + particle combinations and identifies PVs. Bolinger (1971) proposed several syntactic tests of PVs, but these tests have potential shortcomings (see Section 2.2). As a result, this study adopted semantic criteria to identify and classify PVs. Details of the semantic criteria are presented in Section 5.1.

The results of the current study are presented in three chapters: Chapter 6 and Chapter 7 are concerned with how L1 users use PVs, and Chapter 8 compares Chinese EFL learners' use of PVs with L1 users. Chapter 6 focuses particularly on the collocation and the progressive form of specific PVs, whereas Chapter 7 is concerned with the grammatical patterns in which these PVs frequently appears. In addition, Chapter 7 also

shows that it is challenging to replace PVs with one-word equivalents because the patterns where PVs and one-word verbs appear are different. Following the two chapters, Chapter 8 compares the different use of PVs in L1 users' and Chinese EFL learners' writing. In Chapter 8, I also examine concordance lines in detail in order to identify common errors made by Chinese learners when they use PVs in writing.

Chapter 9 is the final chapter in which I review the research questions in the light of the findings presented in Chapter 6 to 8 and try to draw a number of conclusions. This chapter also discusses pedagogical implications for teaching PVs in the Chinese EFL context and shows the strengths and limitations of the research. Finally, this chapter ends with some ideas for further studies.

Chapter 2 English Phrasal Verbs

The aim of this chapter is to establish the complexity of English phrasal verbs (PVs). Many researchers have shown their interests in PVs, but they tended to adopt various technical terms to refer to this particular language form, such as 'phrasal verbs' (Bolinger, 1971; Darwin and Gray, 1999; Liao and Fukuya, 2004; Gardner and Davies, 2007, Garnier and Schmitt, 2015), 'verb-particle combinations' (Fraser, 1976), 'verb-particle constructions' (Gries, 1999; Thim, 2012; Wulff and Gries, 2019; Sung, 2020), or 'particle verbs' (Zipp and Bernaisch, 2012). The present study uses the term 'phrasal verb', as it is the most popular term used in previous studies and in the language classroom settings. ESL/EFL learners and language teachers are more familiar with this term because it frequently appears in dictionaries and classroom materials such as textbooks and course books (Rudzka-Ostyn, 2003).

One crucial problem of teaching and learning PVs might be associated with the lack of a standard definition of this language form (Darwin and Gray, 1999; Liao and Fukuya, 2004). To understand the challenge of studying PVs, this chapter firstly discusses the definition of PVs in Section 2.1. The definition, while attempting to be inclusive, fails to establish a clear boundary between PVs and other multi-word verbs (e.g., prepositional verbs). Then, Section 2.2 will focus on the syntactic and semantic methods of identifying PVs. To be more specific, this section discusses various tests designed to identify PVs (e.g. Bolinger, 1971; Biber et al., 1999) and uses corpus evidence to show the potential weaknesses of these tests. Due to the weaknesses of the syntactic and replacement tests, this section highlights the importance of semantic aspects in the identification of PVs. Finally, in Section 2.3, I will focus on the use of PVs by ESL/EFL learners in both spoken and written discourse. This section also

includes discussions on teaching and learning PVs in the ESL/EFL context.

In addition, it should be noted that this chapter will include some corpus examples for illustration. All corpus data in this chapter were taken from the BoE corpus (see Section 4.2 for more details about this corpus). Unlike examples provided by dictionaries or other classroom materials, the BoE corpus collects data from authentic materials (e.g. books, newspapers, spoken texts, etc.), which is able to reflect how specific PVs are used by experienced users in real oral and written communications.

2.1 Definition of English Phrasal Verbs

As a specific type of multi-word verbs, the basic structure of English PVs consists of two components: the verbal component and the non-verbal component. Linguists have long debated the definition of English PVs, but they have not yet reached an agreement on establishing a standard definition for this linguistic form. For example, Biber et al. (1999) defined English PVs as "multi-word units consisting of a verb followed by an adverbial particle" (p.403). However, Moon (1997) described PVs in a more inclusive way, arguing that PVs "are combinations of verbs and adverbial or prepositional particles" (p.45). The main disagreement between the two definitions is associated with the grammatical role of the non-verbal component in a PV (i.e. whether the particle is adverbial or prepositional).

2.1.1 The verbal component

Linguists generally agree that English PVs have to contain a lexical verb[5]. While it is theoretically possible for any lexical verb to be the verbal component in a PV combination, previous studies have observed a tendency for the choice of the verbal component. Moon (1997, p. 45) found that "the verbs […] of Germanic origin" tended to be particularly productive in the formation of English PVs, such as *come, get, go, put* and *take.* This tendency can be supported by a corpus-based study from Gardner and Davis (2007), who investigated the most frequent English PVs in the BNC. Based on the corpus evidence in BNC, Gardner and Davis (2007) identified twenty lexical verbs that formed the majority of PVs in English, including *go, come, take, get, set, turn,* etc. The observed tendency and the corpus evidence suggested that it was not possible for all English verbs to appear in a PV. For example, the so-called stative verbs[6] would not have the potential to be the verbal component in PVs (Fraser, 1976; Rodríguez-Puente, 2019). Perception verbs[7], such as *feel, hear, see,* etc., were also uncommon to be the verbal component of PVs.

It is worth noting that verbs that are prolific in the formation of PVs also have a high frequency in English. However, these high-frequency verbs usually have multiple meanings (i.e. polysemous), especially abstract meanings (Rudzka-Ostyn, 2003), which may not necessarily result in positive outcomes (Alternberg and Granger, 2001). Although ESL/EFL learners were presumed to have early exposure to these high-

[5] Lexical verb is a grammatical class of words. The lexical verb can "either functions alone, in both finite or non-finite forms … or is preceded by auxiliaries" (Downing and Locker, 2002, p. 15).
[6] Stative verbs describe a state or condition rather than an action or process. Stative verbs often refer to thoughts, emotions, relationships, senses, states of being, and measurements. Examples of this type include *have, know, want, hope,* etc., and they rarely form combinations with particles (Fraser, 1976).
[7] Perceptions verbs, also known as sensory verbs, are verbs that describe the process of perceiving or sensing, such as *feel, hear, see, smell, taste, touch,* etc.

frequency verbs, they still struggled with using them in a semantically and grammatically correct way, even at an advanced level of proficiency (Altenberg and Granger, 2001; Sakaba and Okada, 2021). Therefore, it can be argued that learners' difficulties of using high-frequency verbs could negatively influence their use of phrasal verbs.

2.1.2 The non-verbal component

In terms of the non-verbal component, namely the 'particle'[8], researchers also observed certain tendencies in the choice of the non-verbal component in PVs. Particles such as *up, out, off, in, on, down*, etc. are frequently attached to the verbal component and form a PV (Moon, 1997; Gardner and Davis, 2007). However, compared to the verbal component, the non-verbal component seems to be a more controversial area in the study of PVs (Rodríguez-Puente, 2019).

Traditionally, the difference between PVs and other similar multi-word verb structures can be identified by the type of the non-verbal component that follows the verb (Quirk et al., 1985; Biber et al., 1999; Downing and Locke, 2002). This traditional classification tended to be particularly common in grammar books, where multi-word verbs were divided into three major categories: phrasal verbs (e.g. *pick up*), prepositional verbs (e.g. *look at*), and phrasal-prepositional verbs (e.g. *get away with*) (e.g. Quirk et al., 1985; Biber et al., 1999; Downing and Locke, 2002). Using this traditional classification, researchers firmly believed that the verb was followed by an

[8] Particle is a grammatical term that refers to a small word such as prepositions or adverbs. Particles are typically used in combination with a verb to create a multi-word verb. When attached to a verb, they can change the meaning of the verb and create new idiomatic expressions.

adverbial particle in PVs. As opposed to PVs, the verb was followed by a prepositional particle in prepositional verbs, and it was followed by an adverbial particle and a prepositional particle in phrasal-prepositional verbs (Biber et al., 1999; Downing and Locke, 2002). This classification, however, raises a question about when particles (such as *up* and *down*) are considered adverbial and when they are considered prepositional.

To answer this question and distinguish PVs from the other two types of multi-word verbs, especially prepositional verbs, Biber et al. (1999, p.403) explained that the adverbial particle in a PV carried "core spatial or locative meanings" but it was "commonly used with extended meanings". Side (1990) also noted that the meaning of the adverbial particle in a PV was not always synonymous with the corresponding preposition. The following two examples taken from the BoE corpus demonstrate the difference between the adverbial particle and the prepositional particle.

> She had that style and confidence that made a woman ***stand out*** from the crowd. (BoE: BB_cM032356)
>
> I don't want to ***talk about*** it anymore. (BoE: BA_cF022627)

In the first example, *stand out* is a PV in which the verb *stand* is followed by an adverbial particle *out*. The spatial meaning of *out*, i.e. 'away from a central or inner point' (*Oxford English Dictionary Online*, henceforth *OED Online*), seems to be less helpful when understanding this PV. In *stand out*, the non-verbal component seems to be used in a metaphorical way. The particle suggests that the woman who 'had that style and confidence' became unusual and exceptional, being metaphorically away from 'the crowd'. This metaphorical use of *out* demonstrates what Biber et al. (1999) called the 'extended meaning' of the adverbial particle. On the other hand, as the second example illustrates, *talk about* cannot be regarded as a PV because there is no extended meaning

suggested by the non-verbal component *about*. According to the traditional classification, *talk about* should be labelled as a prepositional verb.

While the examples of *stand out* and *talk about* show the distinction between adverbial particles and prepositional particles, the grammatical function of the particle in some verb + particle combinations can be ambiguous (Side, 1990). For example,

> I ***look into*** his eyes to make sure I note every nuance of expression while he answers me. (BoE: BA_SF042743)
>
> I'm ***looking into*** the possibility of finding a relative or neighbour who can pick them up from school and bring them home. (BoE: NB1__050203)

These two examples suggest that the traditional classification tends to be unclear when the particle *into* can function either as an adverb (i.e. *look into the possibility*) or a preposition (*look into his eyes*). To differentiate the functions of *into* in the two examples, researchers suggested that PVs can be identified using syntactic tests, such as the particle movement test, the passivisation test, etc. (e.g. Bolinger, 1971). These syntactic tests will be discussed in detail in Section 2.2.1.

Because of the unclear grammatical function of the particle, many linguists dismissed the rigid classification and tended to regard PVs as multi-word combinations in which the verb was combined with an adverbial or a prepositional particle (e.g. Cowie and Mackin, 1993; Sinclair, 1996; Moon, 1997). Instead of paying attention to the grammatical classification of multi-word verbs, linguists holding this viewpoint focused particularly on the semantic aspects of PVs, discussing the idiomatic or figurative meanings of verb + particle combinations.

For the purpose of teaching PVs to ESL/EFL learners, Celce-Murcia and Larsen-Freeman (1999) proposed a broader view that PVs consisted of "two (or more) parts that function as a single verb" (p. 426). This broader view suggested that Celce-Murcia and Larsen-Freeman (1999) incorporated what Biber et al. (1999) called 'phrasal-prepositional verbs' into the category of PVs. Celce-Murcia and Larsen-Freeman (1999) employed the label, 'three-word phrasal verbs', to refer to combinations such as *put up with* and *look forward to*, because they argued that these three-word combinations "must be learned as a unit" (p. 427). In more recent studies on PVs, Alangari, Jawrska & Laws (2020) also highlighted the importance of considering PVs beyond two words. The so-called phrasal-prepositional verbs accounted for a substantial proportion in academic texts in the discipline of linguistics (Alangari, Jawrska & Laws, 2020). Despite the significance of phrasal-prepositional verbs in written texts, previous studies seemed to pay less attention to how EFL learners use these three-word PVs. Consequently, the current study will investigate phrasal-prepositional verbs that function as single verbs. In this thesis, I refer to phrasal-prepositional verbs as 'three-word PVs', following Celce-Murcia and Larsen-Freeman's (1999) definition. This label indicates that these three-word PVs share similar characteristics with two-word PVs.

As far as the definition of PVs is concerned, although it is generally agreed that PVs are two-word combinations in which a verb is followed by a particle, previous studies have debated on the second component within a PV. The disagreement is mainly concerned with whether the particle is adverbial or prepositional. Linguists including Bolinger (1971) and Biber et al. (1999) firmly argued that the particle was adverbial in a PV, and they suggested using a number of tests to differentiate PVs and prepositional verbs (see Section 2.2).

2.1.3 Typical features of PVs

2.1.3.1 Syntactic features

In addition to the debate about the grammatical nature of the non-verbal component, the complexity of PVs is also reflected in some typical features of this kind of multi-word verbs. The first typical syntactic feature of PVs is transitivity. Linguists usually categorise PVs into two primary types: intransitive and transitive (Quirk et al., 1989; Biber et al., 1999). Similar to one-word intransitive verbs, intransitive PVs do not take a direct object, such as *come on, shut up, hold on,* etc. The corpus example of *stand out* in Section 2.1.1 also exemplifies that *stand out* is an intransitive PV. On the other hand, transitive PVs can take a direct object, such as *point out (the faults), bring up (two children),* etc. (Quirk et al., 1989; Biber et al., 1999). In English, a number of PVs can function as either intransitive or transitive. The following two corpus examples from the BoE corpus illustrate the intransitive and transitive uses of *TAKE off*, with the first one demonstrating the intransitive use of this PV and the second showing the transitive use.

> If this isn't done, the flights don't ***take off***. (BoE: NB1__050126)
>
> She ***took off*** her backpack and placed it on a chair. (BoE: BA_cF022627)

Another syntactic feature of PVs is particularly applicable to transitive PVs. Researchers found that it was very common for transitive PVs to have a non-contiguous syntactic structure (i.e. the non-verbal component does not follow the verbal component directly) (Quirk et al., 1989; Biber et al., 1999; Gardner and Davis, 2018). For example,

> He ***turned on*** the heater in the car, but it didn't warm him. (BoE:

BA_cM022040)

I ***turn*** the heater ***on*** a little while before I get up. (BoE: NA4__040827)

As the two corpus examples show, *TURN on* is a transitive PV with a flexible structure. The particle could follow the verb contiguously (e.g. '*turn on* the heater'), or non-contiguously (e.g. '*turn* the heater *on*'). The examples also illustrate that the movement of the particle does not affect the meaning of the PV. However, when the verb and the particle are separated by the object, learners may find it difficult to recognize the two components as a single lexical unit (Siyanova and Schmitt, 2007). Since the flexible placement of the non-verbal component is a typical feature of transitive PVs, many researchers suggest that this could be a sufficient criterion to distinguish PVs from other multi-word verbs (Fraser, 1976; Biber et al., 1999). More information about using this feature to identify PVs will be given in Section 2.2.1.

Since it is acceptable for the direct object to separate the verb and the particle in transitive PVs, researchers called this type of PVs separable PVs (Darwin and Gray, 1999; Gardner and Davis, 2018). Gardner and Davis (2018) focused on separable PVs and explored the intervening words between the verb and the particle in the corpus. When a transitive PV took a pronoun as the object, the structure of this PV was predicted to be non-contiguous (Gardner and Davis, 2018). When the verb and the particle were separated by more than one intervening word, as noted by Gardner and Davis (2018), it was challenging to predict the part of speech of intervening words, However, the corpus evidence showed that the verb and the particle were more likely to be separated by noun phrases, such as *pick this one up*, *take the easy way out*, etc. (Gardner and Davis, 2018).

2.1.3.2 Semantic features

Besides syntactic features, the complexity of PVs is also reflected in their polysemous nature. A closer examination of *TAKE off* in the previous examples reveals that the meaning of intransitive *TAKE off* (i.e. 'the flights don't *take off*') is quite different from that of transitive *TAKE off* (i.e. '*took off* her backpack'). In other words, *TAKE off* is a PV with multiple meanings. Similar to those high-frequency verbs with multiple meanings, the polysemous feature of PVs also poses great challenges for ESL/EFL learners (Sinclair, 1996; Siyanova and Schmitt, 2007).

In Darwin and Gray's (1999) definition of PVs, they suggested that PVs generally "[functioned] as a single verb" and the verb and the particle "[gave] up meaning in order to form a new lexical item" (p. 65). This definition indicated another semantic feature of PVs: the meaning of a PV is not equivalent to the combination of the two elements. In other words, the meaning of a PV may not be derived from the meanings of its individual components directly (i.e. non-compositional). PVs with a non-compositional meaning are predicted to be excessively difficult for ESL/EFL learners (Hulstijn and Marchena, 1989; Liao and Fukuya, 2004).

2.1.3.3 Stylistic features

Finally, PVs usually have different distributions across registers (Biber et al., 1999; Liu, 2011). Biber et al. (1999, p. 408) reported that "PVs are used most commonly in fictions and conversations", whereas "they are relatively rare in academic prose". This finding can be supported by Liu (2011), who carried out a cross-register study of the frequently used PVs. According to Liu (2011), the overall frequency of PVs in fiction and spoken English was higher than the frequency in magazines, newspapers, and academic writing.

The findings suggest that PVs tend to be sensitive to register. To achieve a native-like use of PVs, ESL/EFL learners have to employ PVs appropriate to the target register (Siyanova and Schmitt, 2007). For example, overusing PVs in written texts may result in a colloquial and informal tone (Siyanova and Schmitt, 2007; Waibel, 2007).

The uneven distribution of PVs in spoken and written discourse does not suggest that PVs could be dismissed in written discourse, especially in formal writing (cf. Trebits, 2009; Alangari, Jaworska and Laws, 2020; Liu and Myers, 2020). A recent study from Alangari, Jaworska and Laws (2020) showed that PVs such as *go on*, *take up*, and *make up of*, were frequently used in published journal articles in the discipline of Linguistics. Liu and Myers (2020) added *point out*, *carry out* and *bring about* to the list and argued that these three PVs showed a higher frequency in academic writing compared to their frequency in oral speech. Moreover, the meanings of these PVs in academic writing tended to be different from the meanings in spoken or general use (Alangari, Jaworska and Laws, 2020; Liu and Myers, 2020). For example, the authors found that the more frequent meaning of *go on* in academic writing was "to happen, take place" rather than "proceed" (Alangari, Jaworska and Laws, 2020, p.8).

To sum up, although PVs are very common in English, their definition has long been debated by linguists. Issues such as the classification of the non-verbal component, the polysemous nature of PVs, and the flexible structure of separable PVs, etc. make PVs a complex topic, especially in language teaching. ESL/EFL learners may have great difficulties using PVs, and language teachers may also find it challenging to determine which PVs should be prioritized in the classroom. The difficulties of learning and teaching PVs will be discussed in detail in Section 2.3.

2.2 Identifying PVs

As indicated in the previous section (Section 2.1.2), PVs tend to have a fuzzy boundary with other types of multi-word verbs, especially prepositional verbs. To differentiate PVs from other types of multi-word verbs, researchers such as Bolinger (1971) and Biber et al. (1999) have proposed several tests in order to identify PVs effectively. Meanwhile, since the syntactic structure of PVs can be either contiguous or non-contiguous, the search of PVs in the corpus could be troublesome because it is difficult to predict the number of intervening words appearing between the verb and the particle (Gardner and Davis, 2018). Therefore, for the purpose of the current study, it is also important to consider the criteria of identifying PVs in the corpus and rejecting non-PV combinations. This section aims to review the tests proposed in previous studies and evaluate whether these tests are sufficient and reliable to identify PVs.

Tests of PVs proposed in previous studies can be divided into three categories: syntactic, semantic, and phonological (cf. Bolinger, 1971). However, this section will focus on syntactic and semantic tests, ignoring phonological tests. This is because the current research intends to identify PVs in the corpus where the phonological information about specific PVs is unable to display.

2.2.1 Syntactic tests of PVs

Among the various syntactic tests of PVs, researchers often recommend the particle movement test to identify PVs (Bolinger, 1971; Fraser, 1976; Biber et al., 1999). The idea behind this test is associated with one of the typical features of PVs mentioned in

Section 2.1.3 (i.e. transitive PVs usually have a flexible syntactic structure). When the object is a noun phrase, Biber et al. (1999) argue that most transitive PVs allow the flexible movement of the particle, either preceding or following the noun phrase (e.g. *turn on the heater* and *turn the heater on*). Meanwhile, this test is able to eliminate prepositional verbs because it is impossible for a prepositional verb to have a non-contiguous structure (Biber et al., 1999; Darwin and Gray, 1999). For example,

Let's ***focus on*** the facts.

*Let's ***focus*** the facts ***on***. (Darwin and Gray, 1999, p.72)

The particle movement test is therefore claimed to be a better way to distinguish PVs from the instances of multi-word verbs (Fraser, 1976; Biber et al., 1999).

Although this syntactic test is able to show the difference between PVs and prepositional verbs, it has some weaknesses. Firstly, it is mainly designed for transitive PVs and is unable to identify intransitive PVs. Besides, it is useful to identify separable PVs, but useless for inseparable PVs such as *COME across*, *RUN into*, etc. (Celce-Murcia and Larsen-Freeman, 1999; Darwin and Gray, 1999). The following example of *GO through* provides evidence to show the weaknesses of this syntactic test.

We'd all spent many years ***going through*** big changes. (BoE: BA_FF012571)

*We'd all spent many years ***going*** big changes ***through***.

Using the particle movement test, one may claim that *GO through* is a prepositional verb because it does not allow the movement of the particle (e.g. '*going* big changes *through*'). However, whether *GO through* should be regarded as a prepositional verb needs to be examined carefully. The current study would argue that *GO through* is a PV rather than a prepositional verb because the meaning of *GO through* in '*going through* big changes' cannot be predicted by the verbal and non-verbal components (see Chapter

5 for more details about meaning non-compositionality of PVs). As a result, while the particle movement test effectively differentiates transitive PVs and prepositional verbs, it is not a reliable test of all PVs, especially PVs that do not allow the movement of the particle (e.g. *GO through*).

Researchers also used the pronoun test that had a similar theoretical basis with the particle movement test (Bolinger, 1971; Darwin and Gray, 1999). In the particle movement test, the particle can be placed either before or after the noun phrase in a transitive PV. In the pronoun test, however, the particle usually appears after pronouns in transitive PVs (Bolinger, 1971; Darwin and Gray, 1999).

Let's ***take*** them ***on*** in a game of chess.

*Let's ***take on*** them in a game of chess. (Darwin and Gray, 1999, p.73)

As in the particle movement test, the pronoun test is also not applicable to intransitive PVs and there is no corpus evidence showing that this test contributes to identifying inseparable PVs. Also, Bolinger (1971) showed some instances of transitive PVs in which the particle does not precede the object pronouns.

Bring along him and her.

Just ***bring along*** yourself, and we'll do fine. (Bolinger, 1971, p.40)

As the above two examples illustrate, when transitive PVs take reflexive pronouns or multiple pronouns as the direct object, the particle is usually placed after the pronouns (Bolinger, 1971).

The third syntactic test proposed by Bolinger (1971) is that PVs can be identified via the passivisation test. Again, this test is also tailored for transitive PVs and is unable to

identify intransitive PVs. While it is true that the majority of transitive PVs could form passives, Darwin and Gray (1999) argued that this test was less useful for some instances such as *COME across*.

I ***came across*** some old letters in the attic.

*Some old letters were *come across* in the attic. (Darwin and Gray, 1999, p.71)

Apart from transitive PVs, some verb + preposition combinations are possible to have the passive form. Bolinger (1971, p.7) used *TALK about* to illustrate that this test had "limited usefulness" for some prepositional verbs.

They ***talked about*** you.

You were *talked about***.** (Bolinger, 1971, p.7)

As a result, the passivisation test should not be a satisfactory test for identifying PVs. It fails to purely identify transitive PVs, as some prepositional verbs (e.g. *TALK about*) can form passives. Additionally, some transitive PVs (e.g. *COME across*) do not have a passive form.

To avoid identifying unexpected combinations (i.e. prepositional verbs), Bolinger (1971) suggested using the action nominal test and explained that transitive PVs "should yield an action nominal" (p.8). Bolinger (1971) used the following example to demonstrate how PVs form action nominals:

He ***looked up*** the information. → His looking up of the information. (Bolinger, 1971, p.8)

However, Fraser (1976) found that this test may cause confusion when separable PVs were under examination.

He ***threw*** his dinner ***up***.

→ His throwing up of his dinner.

→ *His throwing of his dinner up. (Fraser, 1976, p.3)

Fraser (1976) therefore refined this test by adding that the verb and the particle of a PV cannot be separated when they formed an action nominal. In other words, this test is not applicable to separable PVs with the non-contiguous form. Another weakness of this test was observed by Darwin and Gray (1999), who argued that some transitive PVs did not form action nominals:

I *came across* an old photograph.

*the coming across of an old photograph (Darwin and Gray, 1999, p.72).

Last but not least, Bolinger (1971) introduced the adverb insertion test and argued that adverbs cannot be inserted between the verb and the particle in both transitive and intransitive PVs.

They ***turn up*** regularly on German television. (BoE: JC___96__33)

*They ***turn*** regularly ***up*** on German television.

While this test seems to work with adverbs ending with the suffix *-ly*, some adverbs (e.g. *all, right,* etc.) may violate this test, as they could appear between the components (Fraser, 1976; Quirk et al., 1985).

I'll ***look*** the information right ***up***. (Fraser, 1976, p. 25)

Also, this violation can be supported by corpus evidence. When a PV is separated by two intervening words, Gardner and Davis (2018, p.206) found that "it all" is the most frequent intervening combination in the corpus, as in "*take it all in*", "*figure it all out*", etc.

To sum up, Darwin and Gray (1999) commented that these syntactic tests reflected the various characteristics of PVs but "each [admitted] noteworthy exceptions" (p.75). Among all syntactic tests, most researchers recommended the particle movement test (e.g. Bolinger, 1971; Fraser, 1976). Biber et al. (1999, p.404) believed this test was "more important for combinations with a following noun phrase". However, the particle movement test is particularly tailored for transitive PVs; it cannot identify intransitive PVs sufficiently. The evaluation of other syntactic tests (e.g. the pronoun test and the passivisation test) also suggests that they are not 100% reliable and do not always identify intransitive PVs. Another major weakness of syntactic tests is associated with the identification of inseparable PVs (Celce-Murcia and Larsen-Freeman, 1999; Darwin and Gray, 1999). Most syntactic tests are not very successful at identifying inseparable PVs such as *GO through*, *COME across*, etc. Due to these weaknesses, syntactic tests may ignore potential PVs and provide an unreliable list of English PVs. Therefore, this study will not use syntactic tests to identify PVs in the corpus.

2.2.2 The replacement test of PVs

In terms of the semantic tests of PVs, Bolinger (1971) highlighted it was common to try the replacement test, that is, whether a verb + particle combination can be replaced with a one-word verb. If the combination has a one-word verb replacement, then it is a PV (Bolinger, 1971). The following example shows that *BLOW up* is a PV that can be replaced with the single verb *explode* and the replacement does not affect the meaning of the sentence.

> The public house next to the cinema had just ***exploded***. Another bomb in another nearby pub ***blew up*** a few minutes later. (BoE: BB_Em012020)

> The public house next to the cinema had just ***blown up***. Another bomb in another nearby pub ***exploded*** a few minutes later.

In addition to *BLOW up* and *explode*, Quirk et al. (1985, p.1152) provided more examples of such replacement, including "*give in* and *surrender*", "*catch on* and *understand*".

This replacement test seems to be based on the idea that PVs should be regarded as a single lexical unit in which the meaning of this unit is not equivalent to the combination of the verbal and the non-verbal component. To be more specific, it is not possible to replace *blow* with *explode* alone ('*the nearby pub *exploded up*'). It is also impossible to replace *blow upward* with *explode* ('*the public house had just *blown upward*'), even though *up* and *upward* have a similar directional meaning. This raises the question of how the meaning of *BLOW up* is generated if it cannot be derived from its individual components. Since the focus of the current chapter is to review relevant literature on PVs, I will elaborate on this issue with corpus evidence in Chapter 5.

The replacement test has noticeable weaknesses. Not all PVs have an equivalent one-word alternative and using the replacement test alone may cause problems (Cornell, 1985; Darwin and Gray, 1999; Siyanova and Schmitt, 2007). In addition to one-word verb equivalents, Cornell (1985) noted that the equivalents could be other PVs that were congruent in meaning, as in the pair "*put in for* and *apply for*" (p.274). Darwin and Gray (1999) illustrated that *TAKE over* was a PV that was hard to be replaced with one-word verb equivalents or equivalent PVs. Here is a corpus example showing the challenge of replacing *TAKE over*.

> I was desperate for someone to ***take over*** the finances and administration of Film Gate. (BoE: BA_WM042740)

Based on the contextual information around *TAKE over* shown in the given example, it seems to be difficult to find a one-word verb or a PV that has a similar meaning with *TAKE over*. A possible interpretation of this PVs in the context could be 'gaining the power and taking the responsibility for'.

Another problem of using replacement test is the different registers in which PVs and one-word verbs may occur (Side, 1990; Siyanova and Schmitt, 2007). For instance, although it is grammatically and semantically correct to replace *get back* with *return* in "I will call you when I *get back*/*return* from my trip" (p.121), Siyanova and Schmitt (2007) suggested that there may be subtle differences between the PV and the individual verb. The authors pointed out that the PV (i.e. *get back*) was more likely to appear in spoken and informal English, whereas the one-word verb (i.e. *return*) tended to be used in written discourse such as academic writing or official reports (Siyanova and Schmitt, 2007).

As a result, one strength of the replacement test is that it indicates the meaning of a PV cannot be derived from the meaning of the two components. However, this test is not always reliable, as not all PVs are able to have a one-word equivalent whose meaning is completely synonymous with the PV. It is also true that many PVs and their one-word equivalents cannot replace each other since they differ considerably in terms of style and register.

2.2.3 Identifying PVs in corpus-based studies

So far, this section has presented several syntactic and semantic tests of PVs and

discussed the strengths and weaknesses of each test. Due to the weaknesses of each test, researchers have not yet reached an agreement on which verb + particle combinations should be classified as PVs. Moreover, in much of the work cited in Sections 2.2.1 and 2.2.2 (e.g. Bolinger, 1971; Fraser, 1976), the identification of PVs was largely dependent on researchers' intuition. They usually began with a syntactic or semantic feature of PVs and then provided examples of verb + particle combinations that exemplified at least one of these typical features. However, given that the current study is based on corpus, this top-down approach seems hard to implement. This is because the computer cannot do the tests; it can only retrieve instances of verb + particle combinations.

Darwin and Gray (1999) also raised concerns about the top-down identification of PVs and proposed an alternative approach. The authors called the alternative approach a 'throw-it-out approach', which "[considers] all verb + particle combinations to be potential PVs until they can be proven otherwise" (Darwin and Gray, 1999, p. 76). They believed that throwing out (i.e. a non-PV combination) seemed to be easier than throwing in (i.e. a PV), especially for ESL learners (Darwin and Gray, 1999). This throw-it-out approach has some advantages. Firstly, compared with the conventional top-down approach to PVs, Darwin and Gray's (1999) method is bottom-up, allowing all potential verb + particle combinations to be examined. Also, it could greatly reduce the possibility of including combinations that do not function as PVs (e.g. prepositional verbs) and avoids missing certain instances of PVs (e.g. intransitive PVs) (Darwin and Gray, 1999). Although Darwin and Gray (1999) did not use a corpus-based approach, their idea of throw-it-out inspired the identification of PVs in the corpus (see Section 4.3). Compared to the top-down approach, Darwin and Gray's (1999) method did not

begin with a prescribed feature of PVs; instead, they advised to identify all verb + particle combinations and then eliminate those that did not pass the syntactic or semantic tests.

While the throw-it-out approach reduces the possibility of including non-PV combinations, ESL/EFL learners and language teachers still need to try every test in order to identify PVs and exclude non-PV combinations. Considering the potential weaknesses of syntactic and the replacement tests, learners may feel confused to make the decision. Therefore, the current research intends to propose semantic criteria that may alleviate such confusion caused by syntactic and the replacement tests. The semantic criteria evaluate how the meaning of a PV is constructed and conveyed. Instead of analysing the grammatical function of the non-verbal component (i.e. whether it is adverbial or prepositional), this study draws attention to whether the verbal and non-verbal components create a meaning that differs from the meanings of the individual components. Further details about the semantic criteria will be discussed in Section 5.1.

2.3 Previous studies on teaching and learning PVs

2.3.1 The use of PVs by ESL/EFL learners

A number of studies have investigated the use of PVs by ESL/EFL learners. Generally, these studies found that learners tended to underuse PVs, especially those whose L1 does not contain PVs (e.g. Dagut and Laufer, 1985; Liao and Fukuya, 2004; Chen, 2013; Gilquin, 2015; Sung, 2020). While most studies concluded that L2 learners tended to underuse or avoid PVs, a small number of studies found that PVs were possible to be

overused by L2 learners (e.g. Waibel, 2007). Waibel (2007) compared the use of PVs in LOCNESS (Louvain Corpus of Native English Essays) and G-ICLE (the sub-corpus of German L2 learners in the International Corpus of Learner English). The results showed that the percentage of PVs in the G-ICLE was 24.6% higher than the percentage of PVs in LOCNESS (Waibel, 2007). A possible reason for the overuse by German L2 learners was the existence of "a semantically similar verb type in German" (Waibel, 2007, p.85). However, Chinese is a language that does not contain PVs (Lu and Sun, 2017). Given the lack of PVs in Chinese, the current study speculates that PVs are underused by Chinese EFL learners. They are assumed to avoid using PVs in spoken and written discourse. This section primarily focuses on the underuse of PVs and potential factors leading to this phenomenon.

2.3.1.1 Underuse of PVs

Previous studies showed that underuse seemed to be a common behaviour among learners with different L1 backgrounds and various proficiency levels. For example, Gilquin (2015) provided quantitative evidence to show the underuse of PVs by French-speaking EFL learners. The normalised frequency of PVs in the non-native learner corpus was lower than that in native learner corpus (Gilquin, 2015). In addition to the quantitative evidence, another piece of evidence supporting learners' underuse of PVs was their strong tendency to avoid using PVs when there were one-word equivalents available (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004; Schmitt and Siyanova, 2007).

The absence of PVs in learners' native language has been widely acknowledged as an important factor of avoidance (Dagut and Laufer, 1985; Laufer and Eliasson, 1993; Liao

and Fukuya, 2004; Waibel, 2007; Gilquin, 2015; Sung, 2020). Dagut and Laufer (1985) argued that Hebrew-speaking learners avoided using PVs due to the lack of PVs in learners' L1. Waibel (2007) also reported that Italian learners used fewer PVs in written texts than native students because "Italian […] does not have a phrasal-verb equivalent" (p. 160).

The findings from Chen's (2013) study do not support those of Dagut and Laufer (1985) and Waibel (2007). Chen (2013) found that the overall frequency of PVs in the Chinese learner corpus did not indicate a significant difference from the native novice writers. Since no statistically significant difference was found in Chinese learners' writing, Chen (2013) argued that Chinese learners were able to employ PVs in writing, despite this specific phrasal combination being absent from their mother tongue. This result was probably related to the proficiency level of Chinese learners in Chen's study. Chen (2013) compiled a learner corpus consisting of argumentative essays written by upper-intermediate Chinese learners of English. These students were expected to have extensive exposure to English, as all of them had passed the TEM-4[9] test. According to Chen (2013), passing TEM-4 was roughly equivalent to a score of 6.5 in the IELTS or 550 in the TOEFL. Therefore, despite PVs being absent from their native language, learners' proficiency level and exposure to English could provide underlying evidence to support Chen's argument.

When learners had access to PVs in their L1 (e.g. Dutch), Hulstijn and Marchena (1989) still observed a similar tendency of avoidance. In other words, learners' avoidance of

[9] TEM-4 test is also known as Test for English Majors Grade Four. It is a national test to assess the English proficiency of university students majoring in English (Jin and Fan, 2011).

PVs may not be solely due to the lack of PV constructions in their mother tongue. Hulstijn and Marchena (1989) reported that the semantic complexity of PVs also impeded learners' use of PVs, because Dutch ESL learners tended to avoid those PVs whose meanings were highly idiomatic (e.g. *give up*, *break out*, etc.).

In addition to linguistic factors that were associated with learners' avoidance, researchers also investigated whether the underuse of PVs was related to other factors such as learners' proficiency level (Liao and Fukuya, 2004; Becker, 2014; Wei, 2021) and their language exposure (Chen, 2007; Waibel, 2007; Schmitt and Redwood, 2011; Gilquin, 2023). Liao and Fukuya (2004) found that advanced learners did not exhibit this avoidance behaviour. This finding seemed to agree with the findings from Chen (2013) who also argued that the avoidance behaviour was not evident among advance learners. However, Liao and Fukuya (2004) observed that intermediate Chinese learners showed a strong tendency to avoid using PVs, especially figurative PVs whose meanings cannot be predicted by the meanings of components. In addition, Wei (2021) investigated the use of PVs in spoken language and discovered that learners at higher proficiency level used more PVs than lower-level learners in oral communications. Based on this finding, Wei (2021) argued that Chinese learners tended to use more PVs in oral communications as their level of English proficiency improved.

Finally, learners' everyday exposure to English, such as extensive reading and watching English-language films, appeared to contribute to the improvement of their knowledge of PVs (Schmitt and Redwood, 2011). This argument was supported by Gilquin (2023) who conducted a corpus-based study on the use of PVs with the particle *up* by ESL and EFL learners. ESL learners, who were expected to have a higher degree of exposure,

tended to be more capable of producing native-like use of PVs than EFL learners, whose exposure to English was expected to be limited (Gilquin, 2023).

2.3.1.2 Choices between PVs and one-word verbs

The discussion of learners' avoidance of PVs was often accompanied by the question of what ESL/EFL learners chose instead of PVs (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004; Schmitt and Siyanova, 2007; Waibel, 2007; Sung, 2020). Previous studies concluded that learners, especially those with lower level of English proficiency, preferred to use one-word verbs rather than PVs (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004). One of the factors that resulted in learners' preference for one-word verbs was the semantic complexity of PVs (Hulstijn and Marchena, 1989). Compared with PVs that often have an idiomatic meaning, Hulstijn and Marchena (1989, p.249) stated that "learners may have adopted a play-it-safe strategy […] by choosing the one-word verbs with its more general meaning".

Before corpora were widely available for studies on the use of PVs, conclusions about learners' preference were mainly based on the results of tests such as multiple-choice tests and translation tests (cf. Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004). It could be argued that these conclusions about learners' preference for one-word verbs lacked statistical evidence. Moreover, Siyanova and Schmitt (2007) pointed out that the challenge was not merely about replacing the PV with an individual verb that conveyed the correct meaning. For learners, the more challenging issue was whether they were able to choose the verb that appeared in the appropriate register (Siyanova and Schmitt, 2007). As a result, corpora could be a useful

source for comparing the frequency of PVs and one-word equivalents in learner language and evaluating whether learners were able to make appropriate choices between PVs and one-word equivalents.

Based on the results from the learner corpus, Siyanova and Schmitt (2007) found that one-word verbs were more frequent than multi-word verbs in both written and spoken discourse. Since multi-word verbs tended to be more appropriate in colloquial and informal speech, learners' underuse of these verbs may result in unnaturalness of their spoken language (Siyanova and Schmitt, 2007). The stylistic difficulties were not only evident in learners who underused PVs, but also true in learners who overused PVs in their written essays. As Waibel (2007) noted, German L2 learners showed a tendency to overuse PVs, which made their essays more colloquial and informal compared to essays written by native students. Waibel (2007) concluded that German L2 learners tended to be less sensitive about the stylistic differences between spoken and written registers (Waibel, 2007).

Previous studies have argued that EFL/ESL learners prefer one-word verbs to PVs and pointed out that this may cause stylistic problems. However, researchers seemed to ignore the semantic nuances and grammatical differences between PVs and their one-word equivalents. When ESL/EFL learners made a choice between a PV and its one-word equivalent, previous studies did not elaborate on whether learners were aware of the semantic and grammatical differences between PVs and one-word alternatives. Therefore, this study aims to fill this gap by examining whether learners are able to make a semantically and grammatically appropriate replacement between PVs and one-word verbs.

2.3.2 Teaching and learning PVs in ESL/EFL context

As discussed in Section 2.3.1, learners' difficulties with using PVs are associated with both linguistic (i.e. L1-L2 difference, semantic complexity of PVs) and non-linguistic factors (i.e. learners' proficiency level, the degree of exposure to L2). Since the current study is more concerned with the linguistics factors that may influence learners' use of PVs, this section mainly focuses on linguistics factors. Because of learners' avoidance of PVs, researchers and language practitioners have been finding effective ways that can enhance learners' knowledge of PVs. Some researchers attempted to create frequency-based lists that included the most important PVs (e.g. Gardner and Davis, 2007; Liu, 2011; Martinez and Schmitt, 2012) and the key meanings of the frequent PVs (e.g. Granier and Schmitt, 2015; Liu and Myers, 2020). Besides wordlists, others suggested using cognitive methods to address learners' difficulties (cf. Yasuda, 2010; Lu and Sun, 2017; Lindstromberg, 2022). Since the current study investigates PVs through a corpus-based approach, this section mainly focuses on the frequency-based lists of the most important PVs.

Bolinger (1971) stated that listing PVs could be a strategy to define PVs. Many learner dictionaries use this strategy and list PVs under the entry of verbs, such as the *Oxford Advanced Learner's Dictionary* and *Longman English Dictionary*. However, the listing strategy has potential shortcomings, one of which is that PVs are highly prolific multi-word combinations in English (Bolinger, 1971). New PVs are constantly being produced, and the number of existing PVs continues to increase (Riguel, 2014). The creativity of new PVs suggests that listing PVs might be more feasible if the list is based

on the non-verbal component rather than the verbal component (Bolinger, 1971; Side, 1990). The variety of particles used as a part of PVs is far less than the variety of verbs (Moon, 1997; Biber et al., 1999; Gardner and Davis, 2007). Bolinger (1971) notes that another shortcoming of listing was associated with the difference between British English and American English. Liu (2011) compared the most frequent PVs in BNC and the Corpus of Contemporary American (COCA), two famous corpora representing British English and American English respectively. Liu's findings indicate that the most frequent PVs seem to be similar in both corpora, although users from the two regions may have some different preferences. For instance, American speakers tend to prefer *around* in PVs such as *LOOK around* or *TURN around*, whereas British speakers show a slightly stronger preference for *round* (i.e. *LOOK round* and *TURN round*) (Liu, 2011).

Though creating a full list of PVs seems beneficial, Cornell (1985, p.277) warned that "if no attempt is made to select and grade phrasal verbs", the production of these lists may be "uncomfortably similar […] to the opening of the Pandora's box". It may not be "truly meaningful" if the list of PVs is too long (Liu, 2011, p.667). From the learners' point of view, it is essential to have a list with an explicit focus on the most important and useful PVs (Cornell, 1985; Gardner and Davies, 2007; Garnier and Schmitt, 2015; Liu and Myers, 2020).

Learners' needs therefore raise a question of which PVs should be selected and included in the list. To answer this question, Nation and Waring (1997) highlighted the important role of frequency and believed that learners gained benefit from their vocabulary learning effort because they would meet high-frequency words very often. Cornell (1985) supported the idea of selecting PVs with high frequency and suggested that

corpus should be a valuable source for showing the frequency of PVs and their various meanings. Schmitt and Redwood (2011) also observed a positive correlation between frequency and learner's knowledge of PVs.

Due to the predominant role of frequency in the learning of PVs, a number of corpus-based studies on frequent PVs have been conducted. Among these studies, one of the most famous was from Gardner and Davis (2007), who investigated frequent PVs in the BNC. Gardner and Davies (2007) found that the majority of frequent PVs in the BNC were constructed by 20 lexical verbs (e.g. *go, get, make,* etc.) and 8 particles (e.g. *out, up, down,* etc.). Through a more detailed examination, they discovered that the 100 most frequent PVs accounted for more than half of PV occurrences in the BNC. This list of the 100 frequent PVs should be considered a valuable compilation of high-frequency PVs in English (Garnier and Schmitt, 2015). It was suggested that learners should have extensive exposure to those frequent PVs (Gardner and Davis, 2007).

Despite the value of the list of 100 frequent PVs proposed by Gardner and Davies (2007), Liu (2011) noted that the 100 frequent PVs tended to represent British English rather than other varieties of English, as Gardner and Davies (2007) only used the BNC to investigate frequencies of PVs. Therefore, Liu (2011) conducted a study comparing Gardner and Davies' list with a list that might better represent American English (see Biber et al., 1999 for details about this list). Liu (2011) observed a large amount of overlap between these two lists, with only 4 out of 31 PVs from Biber et al.'s (1999) list being absent from Gardner and Davies' (2007) list. Based on this finding, Liu (2011) provided a list of 152 most frequent PVs in both American and British English.

The two comprehensive lists from Gardner and Davis (2007) and Liu (2011) clearly show the frequent PVs that ESL/EFL learners should be exposed to. However, to enhance learners' knowledge of PVs, Garnier and Schmitt (2015) argued that these two lists could be further developed by indicating the meanings that are more important for learners to know. From Garnier and Schmitt's (2015) perspective, the most frequent meanings of these PVs also made significant contribution to the PV pedagogy. To draw learners and language practitioners' attention to the polysemous feature of frequent PVs, Garnier and Schmitt (2015) proposed a new list of PVs for pedagogical purposes and they called it the PHaVE list. This list included the 150 most frequent PVs in COCA, along with the most frequent meanings for each PV. Garnier and Schmitt's list advised language practitioners to address polysemous PVs in a systematic way, prioritising the most frequent meanings (i.e. the most important) in language instruction and assessment.

Although understanding the most frequent meanings was generally more important for ESL/EFL learners, Garnier and Schmitt (2015) acknowledged that the infrequent meanings of those PVs should also receive explicit attention. Meanwhile, since their list of most frequent PVs was based on the COCA and did not compare how frequently the different meanings occur in different registers, the most frequent meaning of a PV may not be representative across regions and registers. For example, in the PHaVE list, the most frequent meaning of *LOOK up* is "raising one's eyes" (Garnier and Schmitt, 2015, p.657), which seems neutral with respect to the register. However, when *LOOK up* is used in the following example in written discourse, this most frequent meaning fails to make sense (Liu and Myers, 2020).

They use so many big words that I have to ***look up*** in the dictionary. (BoE:

BACXM992664)

In this context, the interpretation of *LOOK up* is not associated with raising eyes. However, the PHaVE list paid more attention to the frequent meanings, taking less account of the peripheral meanings of frequent PVs.

As a result, this study attempts to fill the gap by considering both frequent and infrequent meanings of frequent PVs, investigating learners' knowledge of these meanings and comparing learners' use of frequent PVs with L1 users' use in written discourse.

2.4 Summary

This chapter has reviewed the literature of English PVs and established the complexity of PVs. PVs tend to be challenging for ESL/EFL learners and language teachers for several reasons, including the lack of a standard definition, the various features of PVs, and the difficulties of distinguishing between PVs and other multi-word verbs. Due to the weaknesses of syntactic tests of PVs, I decided to use semantic criteria to identify PVs in the corpus (details about the semantic criteria will be provided in Section 5.1).

Previous studies on teaching and learning PVs in the ESL/EFL context have shown that learners have a strong tendency to underuse or avoid using PVs, especially those whose L1 does not contain such phrasal combinations (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004; Waibel, 2007; Gilquin, 2015; Sung, 2020). In order to enhance learners' knowledge of PVs, many researchers argued that listing could be a useful strategy, and they created usage-based lists of frequent PVs for

teaching and learning PVs (e.g. Gardner and Davis, 2007; Liu, 2011; Granier and Schmitt, 2015; Liu and Myers, 2020).

Despite the extensive research on teaching and learning PVs in the ESL/EFL context, I found that some areas remained under-explored. One of the gaps is associated with learners' choice between PVs and one-word verbs. Learners are found to prefer one-word verbs, but many PVs cannot be directly replaced with one-word verb equivalents (Cornell, 1985; Darwin and Gray, 1999; Siyanova and Schmitt, 2007). The current study aims to investigate the replacement of PVs with one-word equivalents in Chinese EFL learners' writing, analysing whether learners are able to make semantically and grammatically appropriate replacement. The other gap is concerned with the meanings of frequent PVs. Previous studies have largely focused on the frequent meanings of these PVs, overlooking the peripheral meanings. This study tries to fill this gap by investigating the lexico-grammatical patterns associated with PVs and examining how Chinese EFL learners use frequent PVs in their writings.

Chapter 3 Understanding multi-word combinations in context: collocation, grammatical patterning, and verb tense forms

As we have seen from the previous chapter, PVs usually function as meaning units, but they often have multiple meanings depending on the context (Section 2.1.3). By raising the question, 'what does *set* mean [in *set in*]?' (p.67), Sinclair (1991) suggests that PVs are a special type of multi-word units that are difficult to isolate semantically. Knowing the meaning of the particle (i.e. *in*) is not enough; the interpretation of *set in* has to be based on the context (Sinclair, 1991). In other words, understanding the meaning of a PV depends on the lexical or grammatical patterns in which it appears, or a combination of both (i.e. lexico-grammatical patterns). However, providing a comprehensive account of these patterns would require a book-length work due to the vast variety and complexity of these patterns (Hunston & Francis, 2000). The analysis of lexico-grammatical patterns involves numerous word combinations and syntactic structures, including collocation (Sinclair 1991), fixed expressions and idioms (Moon, 1998), formulaic sequence (Wray, 2002), lexical bundles (Biber, Conrad, and Cortes, 2004), pattern grammar (Hunston & Francis, 2000), etc. Although these terms may broadly refer to multi-word combinations, each would have its distinct nuances and discourse functions. For example, collocation is often restricted to two-word combinations (e.g. Stubbs, 2001), whereas lexical bundles usually refer to longer, uninterrupted word sequences that frequently appear in certain registers or genres (e.g. Hyland, 2008). In order to limit the scope of the current study, I will focus specifically on two aspects of lexico-grammatical patterning, collocation and pattern grammar, and provide a systemic overview of these two aspects in this chapter.

In Section 3.1, two different definitions of the term 'collocation' are discussed. By

reviewing the two major definitions, this section also explains the way that collocation was addressed in the current research. Section 3.2 focuses on pattern grammar and investigates how this approach contributes to understanding the relationship between lexical items and grammatical frameworks in which they commonly appear. Finally in Section 3.3, I will briefly discuss the influence of tense forms (particularly the continuous form) on the semantic meaning of verbs. Although tense form is not directly related to the lexico-grammatical patterning discussed in Sections 3.1 and 3.2, it is still relevant, the present study examines PVs whose grammatical function is similar to individual verbs in the language.

3.1 Definitions of collocation

3.1.1 Overview

As multi-word combinations, Moon (1997, p.63) states that "collocation is a very important principle underlying the structure of language". However, collocation differs greatly in various aspects, such as the number of words involved, the type of words that co-occurs (i.e. collocation or colligation[10]), and the fixedness of the form (Nation, 2013). The various types of collocation make it challenging to establish a comprehensive definition for this term (Cowie, 1988; Moon, 1997; Nation, 2013; Nesselhauf, 2004). The complexity of collocation is also related to the different criteria that are used to distinguish collocation from other multi-word combinations (cf. Moon, 1998; Biber, Conrad and Cortes, 2004; Nesselhauf, 2004; Laufer and Waldman, 2011).

[10] A narrow and focused definition of colligation is the co-occurrence of lexical items and grammatical structures (Sinclair, 1991). This is different from collocation which usually refers to the co-occurrence of lexical items. For example, *be interested in* is a colligation in which the adjective *interested* colligates with a grammatical word *in*, whereas *heavy rain* is a collocation in which the adjective *heavy* collocates with the noun *rain*.

Among the various definitions of collocation, Cowie (1988) developed the phraseological view on collocation by situating multi-word combinations on a continuum, with the most flexible and transparent combinations at one end and the fixed and opaque ones at the other. Cowie's approach is deeply influenced by Russian phraseology (Granger and Paquot, 2008), so it is often called the phraseological approach (Nesselhauf, 2004; Durrant and Mathews-Aydinli, 2011) or the lexical approach (Walker, 2011a) in previous studies of collocation. Another widespread definition of collocation is based on Sinclair's idiom principle, which views collocation as the co-occurrence of words within a certain distance. This approach is also referred to as the neo-Firthian approach (Howarth, 1996; McEnery and Hardie, 2012), the empirical approach (Evert, 2009), or the frequency-based approach (Nesselhauf, 2004; Durrant and Mathews-Aydinli, 2011; Walker, 2011a). The current study employs the label 'frequency-based' because a key characteristic of Sinclair's approach is its emphasis on frequency as a crucial indicator of collocation.

The term 'collocation' is also used in distinctive ways in Cowie's phraseological approach compared with Sinclair's frequency-based approach (Liu, 2010; Walker, 2011b). The phraseological approach accepts the countable use of the term because this approach regards collocations as "[combinations] of two or more words which occur together or in close proximity to each other" (Walker, 2011b, p.292). For example, *make a decision* is a collocation and *take a break* is another. However, the term is often treated as an uncountable noun in the frequency-based approach (Liu, 2010; Walker, 2011b). This is because Sinclair and researchers who follow the frequency-based approach regard collocation as a phenomenon or a tendency for certain words to co-occur at a

short distance (see Sinclair's definition of collocation in Section 3.1.3). To be more specific, the words that frequently co-occur do not necessarily "form a habitual combination as a lexical unit" (Liu, 2010, p.5). For example, *toy* and *children* frequently co-occur with each other (Hunston, 2002), but the co-occurrence should be regarded as a matter of collocation rather than a lexicalised combination with a specific meaning (Liu, 2010).

In addition to the two approaches, there is a third approach which regards collocation as a co-occurrence of words that are psychologically associated with each other. Durrant and Mathews-Aydinli (2011) used the label 'psycholinguistic approach' to refer to this third approach. This psycholinguistic approach suggests that users develop a mental repository of words that co-occur frequently, and the co-occurrences are stored holistically through repeated exposure (Wray, 2002; Hoey, 2005). It has some overlapping areas with the phraseological and frequency-based approaches (Durrant, 2014). For example, it aligns with Sinclair's approach by acknowledging that collocation reflects the tendency that certain words occur together more frequently than would be expected by chance (see more details in Section 3.1.3). Hoey (2005) extended this idea by indicating that words are mentally primed not only for collocational use, but also for semantic and grammatical associations. Although I included some supplementary comments made by Hoey (2005), this chapter does not focus on this psycholinguistic approach because it is not a completely new approach to the identification of collocation.

Therefore, the following sub-sections focus primarily on Cowie's phraseological approach (Section 3.1.2) and Sinclair's frequency-based approach (Section 3.1.3), and

then introduce how the current study defines and identifies collocation in the corpus (Section 3.1.4).

3.1.2 The phraseological approach to collocation

Overall, Cowie (1988) categorised word combinations into two broad groups, 'composites' and 'formulae'. Formulae refer to word combinations with a pragmatic or discourse function (Cowie, 1988; Granger, 1998). For example, both fixed expressions such as "*good morning* and *how are you*" and less frozen sentence builders such as "*it can be claimed/assumed that…*" are regarded as formulae (Granger, 1998). In contrast to formulae, composites are combinations whose forms are relatively invariable and whose meanings are relatively unitary (Cowie, 1988). Cowie (1988) indicated that collocations should be considered a member of composites, and the other member of composites was idioms. Under the two broad categories, four types of combinations can be identified, with 'free combinations' and 'restricted collocations' falling under one category, and 'figurative idioms' and 'pure idioms' belonging to the other (Cowie, 1981; Howarth, 1996). The diagram in Figure 3.1 was created to visualise the relationships among the four types of combinations. However, it should be noted that the boundaries between these types are not clearly defined (Cowie, 1981; Howarth, 1996). Cowie (1988) presented these four types on a continuum, and thus Figure 3.1 also shows the position of each type along this continuum.

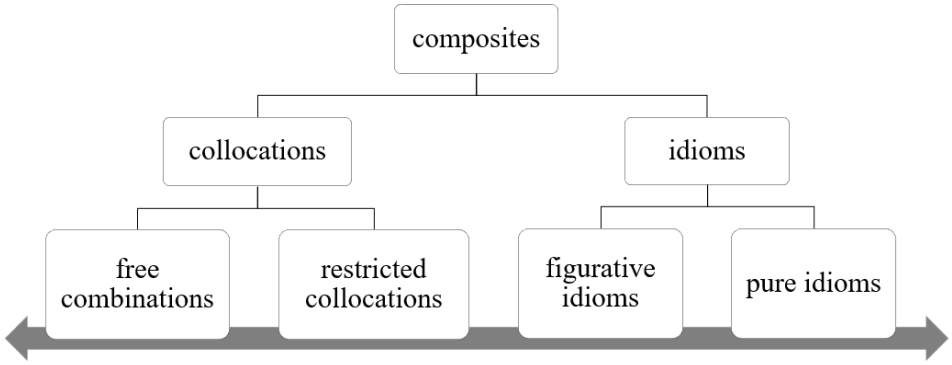


Figure 3.1 Cowie's phraseological continuum of word combinations in English

The classification of the four types of word combinations is based on two primary criteria: semantic transparency and commutability (Cowie, 1981; Howarth, 1996). Semantic transparency describes the extent to which the individual components of the combination have a literal meaning (Howarth, 1996). Commutability, which is also known as restrictedness, refers to "the restrictions on the possible substitution of one element in the [combination] without a consequent alteration in the meaning of the other element" (Howarth, 1996, p.41). In other words, the second criterion is concerned with the extent to which the substitution of an individual component in the combination is restricted by the other.

As Figure 3.1 shows, free combinations, which are also called 'open collocations'[11] by Cowie (1981), appear at one end of the continuum. This type includes word combinations in which each individual component is used in a literal way (Cowie, 1988). In other words, no additional interpretation or metaphor is involved in understanding the meaning of free combinations. Considering the two criteria (i.e. semantic transparency and commutability), *read a book* can be seen as an example of

[11] The two labels, 'free combination' and 'open collocation', are similar in the phraseological approach to collocation, although some researchers do not regard 'open collocations' as collocations (e.g. Howarth, 1996). Howarth (1996) uses 'free combination' and argues that "the term collocation should be reserved for expressions that are to some extent restricted" (p.34).

free combinations because each component in this combination has a literal meaning. One component can be substituted without affecting the meaning of the other (e.g. *buy a book* or *read a letter*). Pure idioms, which are situated on the other end of the continuum model, refer to combinations with a semantically opaque meaning and grammatically fixed form (Cowie, 1988). Many researchers considered *kick the bucket* a satisfactory example of pure idioms, as the meaning of this combination cannot be derived from the literal meanings of the components (Moon, 1997). More specifically, the meanings of *kick* and *bucket* fail to predict the meaning of *kick the bucket* whose meaning is 'to die'. Meanwhile, the form of this combination is fixed, with no possibility for substitution. For instance, if *kick the bucket* is substituted by **kick a bucket*, the substitution is no longer an idiom and completely loses its idiomatic meaning.

Between the two extremes lies what Cowie (1981) calls "restricted collocations" and "figurative idioms". In contrast to free combinations in which each individual component has a literal meaning, restricted collocations are combinations in which "one component is used in its literal meaning, while the other is used in a specialised sense" (Howarth, 1996, p.47). For example, in *make a decision* and *take a break*, the nouns (i.e. *decision* and *break*) preserve their literal senses, whereas the verbs (i.e. *make* and *take*) are delexicalised; that is, both the meanings of *make* and *take* become very weak in the combinations. The verbs, although lose the lexical meanings, are used to support the nouns (i.e. *decision* and *break*), which carry the main meaning of the combinations. The final type on the continuum model is figurative idioms, which are combinations with figurative meanings but still allow for literal interpretations, such as *change gear* in the sense of 'alter one's approach' (Cowie, 1981; Howarth, 1996). Howarth (1996)

noted that figurative idioms may allow the substitution of one component, but such substitution was quite arbitrary and rarely possible. In the case of *change gear,* it seems that the only plausible substitution could be *shift gear*.

Based on the above description of the four types of word combinations, it can be observed that there is a clear distinction between free combinations and pure idioms, as they differ significantly in terms of their semantic transparency and restrictions on substitution (e.g. *read a book* vs. *kick the bucket*). However, the two criteria proposed by Cowie may cause problems when drawing clear distinctions between the four types (Nesselhauf, 2004). Nesselhauf (2004) used the combination *commit a crime* to exemplify the problems of using two criteria at the same time. Using the criterion of semantic transparency, the combination should be classified as a free combination because both *commit* and *crime* were used in their literal senses (Nesselhauf, 2004). However, using the criterion of commutability, this combination cannot be classified as a free combination because it was restricted in its commutability (Nesselhauf, 2004). As a result, Nesselhauf (2004) argued that the criteria of semantic transparency and commutability "[did] not regularly coincide" (p.25) when defining the distinction between collocations and other types of word combinations.

Due to the fuzzy boundaries of collocations in Cowie's continuum, researchers who followed the phraseological approach usually made their own definitions of collocations (e.g. Nesselhauf, 2004; Laufer and Waldman, 2011). For example, Laufer and Waldman (2011) focused on verb-noun collocations and classified three types of verb-noun combinations: free combinations, collocations and idioms. They defined collocations as "habitually occurring lexical combinations" (p.648) using two criteria:

(1) the co-occurrence of components is restricted, and (2) the meaning of the combination is relatively transparent (Laufer and Waldman, 2011). Laufer and Waldman (2011) used the first criterion to differentiate collocations and free combinations because individual components in a free combination can be substituted easily based on the grammar rules, whereas components in a collocation cannot. The second criterion was used to differentiate collocations and idioms because the meaning of a collocation was somewhat more transparent, whereas the meaning of an idiom tended to be opaque and cannot be predicted by the components. Compared with Cowie's phraseological continuum, the major difference in Laufer and Waldman's (2011) classification is the absence of restricted collocations and figurative idioms. It seems that Laufer and Waldman's (2011) focused more on distinguishing collocations from free combinations and idioms, while they ignored the distinction between restricted collocations and figurative idioms. Unfortunately, Laufer and Waldman (2011) did not provide explanations on how they incorporated restricted collocations and figurative idioms into their classification.

Similar to the classification in Laufer and Waldman's (2011) research, Nesselhauf (2004) also examined verb-noun collocations and classified verb-noun combinations into free combinations, collocations and idioms. However, in Nesselhauf's (2004) classification, she explained that figurative idioms were absent because this category was assigned either to collocations or to idioms, "depending on whether the noun in the given sense can be combined freely" (p.33). Nesselhauf (2004) attempted to delimit the boundary between free combinations and collocations by examining the meaning of the verb. The former included verb-noun combinations in which the choice of verbs was not restricted by the noun (e.g. *want/buy a car*), whereas the latter included verb-noun

combinations in which the choice of verbs was arbitrarily restricted by certain nouns (e.g. *shrug one's shoulder* and *make a decision*). In terms of collocations and idioms, Nesselhauf (2004) described that the distinction between these two types was dependent on the noun. For instance, in the idiom *spill the beans*, it was hard to isolate the meaning of the noun, whereas the meanings of *shoulder* and *decision* were preserved in the two collocations (Nesselhauf, 2004).

To sum up, Cowie's phraseological approach to multi-word combinations distinguishes four types of combinations on the basis of two primary linguistic criteria, semantic transparency and restrictions on substitution. Due to the fuzzy boundaries between each type of combinations, Cowie's approach regards multi-word combinations as a continuum, with free combinations (semantically transparent and syntactically flexible) verging on one end and pure idioms (semantically opaque and syntactically fixed) verging on the other. Although there was no clear-cut boundary of collocations on this continuum, researchers who followed this approach tended to regard collocations as word combinations falling in between the two extremes (e.g. Nesselhauf, 2004; Laufer and Waldman, 2011).

3.1.3 The frequency-based approach to collocation

In contrast to Cowie's phraseological approach, which identifies collocations using semantic and syntactic criteria, the frequency-based approach, originated from Sinclair's lexicographic work, is a bottom-up, corpus-driven method that identifies word co-occurrences in the corpus (Granger and Paquot, 2008). As mentioned in Section 3.1.1, the use of collocation is uncountable in the frequency-based approach, because Sinclair (1991) treated collocation as a phenomenon of lexical co-occurrences

rather than a specific category of word combinations. This uncountable use can be reflected by Sinclair's (1991) definition on collocation, that is "the occurrence of two or more words within a short space of each other in a text" (p.170). This definition is influenced by the collaboration of two principles, the open-choice principle and the idiom principle, which can be seen as the theoretical foundations for Sinclair's approach to the co-occurrence of words.

The open-choice principle can be understood as "a 'slot-and-filler' model" (Sinclair, 1991, p.109). It views the text as a collection of slots where random words, as meaning units, are virtually allowed to fill the slot as long as the choice of fillers is grammatically correct (Sinclair, 1991). In other words, this principle regards grammar as the underlying constraint on the choice of words (Barnbrook, 2009). The open-choice principle reflects the syntagmatic feature of multi-word combinations; that is, the linear order in which words appear (Granger and Paquot, 2008). However, Sinclair (1996) argued that the word choice in a given slot was rarely random in real language use. This argument could be supported by two combinations: *strong tea* and *heavy rain*. According to the open-choice principle, it should be possible to have two other combinations: **heavy tea* and **strong rain*, as both of them are grammatically correct. If we search for *strong tea* and **heavy tea* in the corpus, unfortunately, it is hard to find evidence to support the claim that the two adjectives are interchangeable. In the BoE corpus, there are numerous instances showing the adjective *strong* frequently co-occurs with the noun *tea* (see Figure 3.2), whereas there is no corpus evidence showing that L1 users choose *heavy* to precede *tea*.

I made myself a pot of **strong** **tea** and a couple of toasted pikelets
sugar with them, coffee and **strong** **tea** on a regular basis
it is very **strong** black **tea**, which is made with about four spoons of sugar
You drink first. The **tea** was **strong** and bitter,
It was a sticky day and the **tea** was **strong** and milky.

Figure 3.2 The co-occurrence of *strong* and *tea* in the BoE corpus

Similarly, it is not possible to find instances of **strong rain* in the BoE corpus although the linear order looks correct. When *strong* appears in the context of *rain*, as Figure 3.3 demonstrates, this adjective seems to have a closer connection with *wind* or *wave* (i.e. *strong wind* or *strong waves*) rather than *rain*.

a storm comes with **strong** winds and **rains**, the elder brother loses his luck
expected to deliver **strong** waves, **rain** and wind to Mexico
After several days of heavy **rain** and **strong** winds, weather conditions began to
Officials blamed **rain** and **strong** winds for the outages
forecasters were predicting no **rain** but **strong** gusts of wind today.

Figure 3.3 When *strong* appears in the context of *rain* in the BoE corpus

The two examples of *strong tea* and *heavy rain* illustrate that the open-choice principle fails to explain why *heavy* does not co-occur with *tea* and *strong* does not co-occur with *rain* (i.e. the impossibility of **heavy tea* and **strong rain*). Complement to the open-choice principle, Sinclair (1991) introduced the idiom principle, which indicated the restricted choice of words. The idiom principle is described as "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair, 1991, p.110). This principle indicates that experienced language users do not produce *strong tea* and *heavy rain* from scratch; they may store these two expressions as semi-preconstructed combinations (Barnbrook, 2009). It also explains why **heavy tea* and **strong rain* are extremely rare in the BoE corpus. This is because the adjective and the noun co-select each other, constituting a single choice. As a result, Sinclair (1991) stressed that lexis should be considered as important as grammar when describing how meaning arises in a text. Unlike the open-choice principle that reflects the syntagmatic

relations[12] of words, the principle of idiom deals with the paradigmatic relations[13] between words. It indicates that the potential choices of words that can fit into a given slot are not random (Sinclair, 1991; Barnbrook, 2009).

The description of the two principles suggests that the open-choice principle accounts for the creative aspect of language, while the idiom principle accounts for the formulaic aspect. In actual language use, Sinclair (1991) argued that the idiom principle appeared to be the default mode when interpreting most texts, with occasional shifts to the open-choice principle. Collocation, which refers to the co-occurrence of words, illustrates the idiom principle. However, not every co-occurrence of words qualifies as a collocation. To define what qualifies as a collocation, Sinclair (1991) distinguished two types of collocation, "casual collocation" and "significant collocation". Casual collocation refers to non-significant collocation (i.e. the co-occurrence of words is accidental or unexpected) (Barnbrook, 2009). In contrast, significant collocation involves words that co-occur more frequently than would be predicted based on their individual frequencies and the length of the text in which they appear (Sinclair, Jones and Daley, 2004). In Sinclair's research on lexical co-occurrences, as well as in other studies employing this frequency-based approach (e.g. Stubbs, 2001), the term 'collocation' is often referred to as significant collocation that are of statistical significance.

Some technical terms are key to understanding Sinclair's frequency-based approach.

[12] Syntagmatic relations between words can be described as "words that go together in a syntactic structure" (Murphy, 2003, p.8).

[13] In contrast to syntagmatic relations, paradigmatic relations refer to "some sort of paradigm […] that contains members of the same grammatical category that share some […] characteristics in common, but fails to share others" (Murphy, 2003, p.8). For example, words such as *blue, black, red, white* are members of the colour paradigm that can fit into the phrase "a ___ chair" (Murphy, 2003, p.8). However, in another phrase 'a ___ apple', it seems that only *red* can properly fit into the slot. This example illustrates the idiom principle, indicating the restricted choices of words.

The target lexical item under investigation is called ‘node’; words that frequently co-occur with the node word in the co-text is called ‘collocate’ (Sinclair, 1991). While the term ‘context’ may encompass both linguistic and non-linguistic situations in which the language is used, the ‘co-text’ of the node consists of “the other words on either side” (Sinclair, 1991, p.172). As the previous corpus examples in Figure 3.2 show, the node word is *tea*, which is often centred in concordance lines, and *strong* is an adjective collocate of the node in the co-text. According to Sinclair’s (1991) definition, the node and the collocate co-occur at a certain distance, which is often referred to as the ‘span’. The term ‘span’ refers to the number of words that precede and/or follow the node within which collocates are identified (Stubbs, 2001). Intuitively, it might be assumed that the node would have a stronger association with directly adjacent collocates, with the associative link becoming weaker and less consistent as the length between the node and collocates become larger (Clear, 1993). However, the notion of a span indicates that the node and the collocate are not necessarily adjacent to each other and researchers often propose different sizes of the span when studying collocation. For example, Stubbs (1995) investigated the significant collocates of the node *CAUSE* within a span of 3:3 words in the corpus. Sinclair, Jones, and Daley (2004) indicated that the optimal span was usually around four words to the left and right side of the node (i.e. a span of +/- 4 words, or 4:4). Table 3.1 shows the five concordance lines of *tea* in Figure 3.2, but the length of each line has been shortened to display only a span of four words on both sides of the node.

Table 3.1 Words appearing in the 4:4 span of *tea* in the BoE corpus

	N-4	N-3	N-2	N-1	Node	N+1	N+2	N+3	N+4
1	a	pot	of	**strong**	**tea**	and	a	couple	of
2	them	coffee	and	**strong**	**tea**	on	a	regular	basis
3	very	**strong**	and	black	**tea**	which	is	made	with
4	You	drink	first	The	**tea**	was	**strong**	and	bitter
5	sticky	day	and	the	**tea**	was	**strong**	and	milky

Table 3.1 shows that the adjective collocate *strong* appears on both sides of the node *tea* within a span of 3:3 words. The table also shows labels used for each position around the node, such as N-3, N+2, etc. For example, the collocate *strong* appears at the N-1 position (i.e. one word to the left of the node) in lines 1 and 2, whereas it appears at the N+2 position (i.e. two words to the right of the node) in lines 4 and 5. This thesis will also use these labels to show the position of collocates.

Based on the above discussion on the frequency-based approach to collocation, Sinclair (1991; 1996) emphasized that words do not randomly occur with other words. Beyond collocation, Sinclair (1996) observed colligation, which referred to the habitual co-occurrence of words with grammatical choices. Hunston (2001) states that studying colligation is helpful because it draws attention to the repeated grammatical behaviour of words. Moreover, studying colligation also helps to understand the semantic and pragmatic functions (i.e. semantic prosody) of a pattern (Sinclair, 1996). The colligational pattern tends to be controlled by semantic preference (Sinclair, 1996). For example, the phrase *the naked eye* frequently co-occurs with grammatical words such as *with*, *to*, etc. (i.e. prepositions), in the pattern "visibility + preposition + the + naked + eye" (p.87). This pattern suggests that the semantic preference of *the naked eye* is verbs or adjectives conveying the meaning of visibility, such as *see* or *visible.* Moreover, this meaningful pattern associated with *the naked eye* can be interpreted as "expressions

of some difficulty" because it has a semantic prosody of difficulty (Sinclair, 1996, p.88). The observation of semantic prosody in collocation reflects Sinclair's (1991) argument with the idiom principle, which is "words and phrases show a tendency to occur in a certain semantic environment" (p.112). Louw (1993, p.157) agreed with this semantic consistency observed in collocation and defined semantic prosody as "a consistent aura of meaning with which a form is imbued by its collocates". Moreover, Louw (1993) categorised the semantic prosody into negative and positive based on the findings of *utterly* (has a negative semantic prosody) and *build up* (a positive semantic prosody). Stubbs (1995) also found a negative semantic prosody of *CAUSE* because this lemma frequently co-occurred with words such as *anxiety, complaint, trouble*, etc. in the corpus. Although some dictionaries defined *CAUSE* in a neutral way, as in "a cause is something which produces an effect" (p.47), Stubbs (2001) argued that this lemma had a strong tendency to co-occur with negative words and expressions based on corpus evidence. Although there are two categories of the semantic prosody in general, Louw (2000) pointed out that negative semantic prosodies were more frequent than positive ones.

To sum up, in contrast to Cowie's phraseological approach to collocation, Sinclair identifies collocation by examining how frequently two or more words occur together in a certain distance. Based on the idiom principle, Sinclair (1991) argues that words co-select each other, and the habitual co-occurrence reflects the lexical patterning of words. In addition to collocation, Sinclair (1996) also extends that the lexical patterns are meaningful and usually indicate the semantic prosodies of words. The notion of semantic prosody connects the meanings of lexical items with communicative purposes (Stubbs, 2009).

3.1.4 The definition of collocation in the current study

The previous sections have reviewed the two major approaches to collocation, the phraseological approach and the frequency-based approach. They are indeed two different approaches to multi-word combinations. Evert (2009) demonstrated the difference between them with the help of the collocation *good* and *time*. Evert (2009) explained that *good* and *time* were significantly collocated according to the frequency-based approach, whereas *a good time* may not be identified as a collocation by the phraseological approach (Evert, 2009). This was because *a good time* had a compositional meaning (i.e. semantically transparent) and accepted non-restricted substitution (e.g. *a bad time* or *a good day*). Nevertheless, these two approaches are complementary rather than contradictory (Granger and Paquot, 2008; Evert, 2009). Evert (2009) indicated that there was considerable overlap between the two approaches, as many collocations identified through corpus analysis had phraseological characteristics. For example, the significant collocation *strong tea* can reflect lexical restrictedness as the adjective cannot be substituted by other synonymous adjectives (e.g. **heavy tea* or **powerful tea*). Likewise, restricted collocations, such as *make a decision*, whose meaning is not fully transparent is expected to be statistically significant in the corpus.

Many researchers suggested that a major issue in collocation research was the need to define the scope of the study and establish criteria for classifying collocations, either phraseological or statistical, in order to present clearer and more valid results (Nesselhauf, 2004; Walker, 2011b; Nation, 2013; Durrant, 2014). Therefore, the aim of this section is to specify how the current study deals with multi-word combinations.

When investigating the use of collocations by German-speaking learners of English, Nesselhauf (2004) followed the phraseological approach and developed a narrow definition of collocation (Section 3.1.2). Stubbs (1995) used the frequency-based approach to investigate the frequent collocates of *CAUSE*. In this study, the node was the lemma *CAUSE*, including all instances in which the node was used as a verb or a noun. The study captured collocates that co-occurred frequently in a span of 3:3 words. Durrant and Schmitt (2009) also used the frequency-based approach, but they focused particularly on adjacent word pairs such as adjective-noun and noun-noun collocations. Without defining the node word, Durrant and Schmitt (2009) directly retrieved collocations in the corpus using two statistical measures, *t*-score and mutual information. The two measures are the most widely used statistical measures to identify significant collocations and more details about the statistical measures will be presented in Section 4.4.

The current study uses the frequency-based approach to identify nominal collocates that frequently co-occur with PVs. The node words are PVs in which the verbal component was analysed at the lemma level. The distance between the node and the collocate is restricted to a 3:3 span. Compared with the optimal 4:4 span recommended by Sinclair, Hones and Daley (2004), the 3:3 span is more limited but fit into the current study for two reasons. Firstly, the collocational association between the nouns and the node PV tends to be weaker as the distance increases. Secondly, when the verb and the particle of a PV are separated by an intervening item (e.g. *look the word up*), the chance of counting less significant collocates increases, as the most significant nominal collocate tends to appear between the verbal and the non-verbal components.

3.2 Pattern grammar

3.2.1 The definition of pattern grammar

Sinclair (1991, p.91) argues that "there is a strong tendency for sense and syntax to be associated". This argument has been further developed by Hunston and Francis (2000) who introduce 'pattern grammar' to show the association between sense and syntax. In general, the nature of 'pattern' is repetition. Hunston (2010) explains that a noticeable pattern emerges from the repeated use of combinations of words and grammatical structures. Based on the notion of patterns, the repeated occurrences of collocates could be understood as the lexical patterns associated with a node word. In contrast to collocation which particularly describes the lexical patterning of words, pattern grammar, drawing on a large amount of corpus evidence, mainly describes the grammatical patterning of a words. The repeated grammatical patterns associated with a word typically encompass small words (such as prepositions), groups and clauses that appear around the word (Hunston and Francis, 2000; Hunston, 2008).

Briefly, Hunston and Francis (2000) define the pattern of a word as "all the words and structures which are regularly associated with the word, and which contribute to its meaning" (p.37). Hunston and Francis (2000) investigated grammatical patterns in the BoE corpus and provided a number of corpus examples to demonstrate the association between patterns and words. Compared to the traditional descriptive grammar that operates entirely on the syntactic level, this corpus-driven approach to grammar extends the traditional view by taking lexis into consideration (Francis, 1993; Hunston, 2008). The investigation of patterns in the corpus shows the mutual dependence between words and grammatical structures (Francis, 1993; Hunston and Francis, 2000). For

example, when the introductory *it* is used as the object, corpus evidence shows that *find* and *make* are two most frequent verbs that precede *it*, whereas other verbs such as *think* and *consider* are less frequent. Francis (1993) argued that traditional grammar did not indicate such restricted choice of verbs explicitly. Meanwhile, Francis (1993) also observed that the communicative function of the grammatical structure tended to be clearer via corpus investigation. For instance, corpus evidence showed that the choice of adjectives (such as *difficult, hard*, etc.) was also restricted in '*find it* + adjective' and this restricted choice was closely connected with the evaluative meaning of this structure (Francis, 1993).

3.2.2 Two major arguments of pattern grammar

Based on the extensive investigation of patterns in the BoE corpus, Hunston and Francis (2000) present two major arguments for pattern grammar in their book. The two major arguments will be discussed in detail in this section.

Firstly, Hunston and Francis (2000) argue that patterns and lexis are mutually dependent with each other. Each pattern is dependent on a restricted choice of lexical items, and each lexical item occurs relatively frequently with a restricted set of patterns (Hunston and Francis, 2000). For example, by analysing 50 random concordance lines of the verb *explain*, they found that this verb is frequently used in complementation patterns, such as **V n *to* n**[14] or **V wh**[15] (corpus evidence showing the grammatical patterning of *explain*

[14] Hunston and Francis (2000) proposed the way of presenting patterns in their book (p.45). Where a preposition, adverb or other lexical item is part of a pattern, they suggested using italics to indicate that it is a lexical item rather than an open slot for a group of words. In the pattern **V n *to* n**, for example, *to* is a preposition which introduces the indirect object of the verb.

[15] **V wh** is a pattern that shows verbs are followed by clauses beginning with a wh-word (including *what, when, where, which, who, why,* and *how*). The slot for verbs is capitalised (i.e. **V**), which indicates that it is a pattern associated with verbs.

can be found in Hunston and Francis, 2000, p.38). While these two patterns are compatible with *explain*, the verb slot in **V n *to* n** does not accept other verbs such as *come* and *make*. This is an important observation because it indicates that patterns occur with a restricted choice of words (Francis, 1993; Hunston and Francis, 2000). Additionally, a lexical item also occurs in a restricted set of patterns (Hunston and Francis, 2000). This can be illustrated by an adjective, *afraid*. In the BoE corpus, *afraid* frequently occurs in two different patterns, **ADJ *of*** (*be afraid* of) and **ADJ *that*** (*be afraid that*), but it never occurs in the **ADJ n** pattern. More precisely, *afraid* does not have the attributive use because it never occurs before a noun. Hunston and Francis's observation provides corpus evidence demonstrating the interdependence between lexis and grammatical patterns, arguing against the traditional view that lexis and grammar are separate domains in language (Francis, 1993; Groom, 2005; Römer, 2009).

The second argument on pattern grammar is concerned with the association between pattern and meaning (Hunston and Francis, 2000; Groom, 2005). Hunston and Francis (2000) outlined two valuable observations from the large-scale corpus data. One observation was that different meanings of a polysemous word tended to appear in different patterns; the other observation was the tendency for words from the same meaning group (i.e. they share a meaning aspect) to share the same pattern (Francis, 1993; Hunston and Francis, 2000; Groom, 2005).

The previous example of *afraid* can illustrate the first observation. Considering the following corpus examples in Figure 3.4:

John Eglinton that he was not **afraid of** death; all he objected to was
The slow is: If you are **afraid of** deep water, enter the water each day
I've always been a little **afraid of** dogs - till I realised he only wanted to
reserved and perhaps a little **afraid of** making mistakes, the boy's former
been to our home. I was **afraid of** what would happen after their lunch

She was also **afraid that** doing so might make John think
my letter to you. I am **afraid that** I am not prepared to release
were Kells or Picards. I am **afraid that** I am totally unaware of any stories
How old is Chung? I'm **afraid that** I cannot tell you,
stumbled out of his room, **afraid that** it was quite late in the morning

Figure 3.4 Random concordance lines of *be afraid of* and *be afraid that* in the BoE corpus.

As mentioned before, the adjective *afraid* frequently occurs in the **ADJ *of*** and **ADJ *that*** patterns. However, the meaning of this adjective in *be afraid of* is different from that in *be afraid that*. Corpus evidence exhibited in Figure 3.4 shows that the meaning of *afraid* in *be afraid of* is associated with 'being frightened', whereas the meaning is associated with 'apologising for something' when it appears in the pattern *be afraid that* (Hunston and Francis, 2000).

In addition to the example of *afraid*, Hunston and Francis (2000) also investigated the use of the verb *reflect* in the corpus and analysed the meanings of this verb. According to their findings, there are three frequent meanings of *reflect* observed in the corpus (i.e. light-related, mirror-related, thinking-related) and each of them "typically occurs in a particular phraseology, that is, collocating with different types of noun or pronoun [...] and with a different complementation pattern" (Hunston and Francis, 2000, p.255). Table 3.2 summarises their findings and shows how grammatical patterns associated with *reflect* distinguish the meanings of this verb.

Table 3.2 How grammatical patterns distinguish the different meanings of the verb *reflect*, based on the findings from Hunston and Francis (2000)

	Meanings	Patterns	Examples
reflect$_1$	related to light or surfaces	**V n**	*The glass appears to **reflect** light naturally.*
reflect$_2$	mirror-related	***be* V-ed**	*His image was **reflected** many times in the mirror.*
reflect$_3$	thinking-related	**V prep**	*I **reflected** on the child's future.*

As Table 3.2 shows, the light-related meaning of *reflect* (i.e. *reflect*$_1$) occurs more frequently in the **V n** pattern, whereas the thinking-related meaning (i.e. *reflect*$_3$) is more likely to be found in the **V prep** pattern. Different meanings of *reflect* are signalled by different patterns (Hunston and Francis, 2000). In other words, it is not impossible to find that, for example, *reflect*$_1$ is followed by a preposition in the corpus.

In terms of the second observation, the association between pattern and meaning was firstly indicated by Francis (1993), who observed that words from the same meaning group tended to appear in the same pattern. Hunston and Francis (2000) illustrated this tendency with the help of a specific verb pattern, **V *over* n**. They investigated a number of concordance lines in which this pattern appeared and found that the verb slot before the preposition *over* can be filled with a small group of verbs, such as *disagree, fight, quarrel,* and *wrangle* (p.44). These verbs shared a common aspect of meaning, which was associated with 'having an argument or discussion'. Therefore, in the reference book of verb patterns, Francis et al. (1996) included these verbs in the 'argue' group.

Furthermore, Hunston and Francis (2000) observed that the pattern itself was possible to carry a meaning. For instance, they illustrated that the pattern ***there* v-link *something* ADJ *about* n** usually carried an evaluative meaning, evaluating the person or thing indicated in the noun slot following *about* (Hunston and Francis, 2000). Considering

the following examples extracted from the BoE corpus:

> You've always known there was something **odd** about the whole situation surrounding the horse. (BoE: BA_fF042799)
>
> She examined it closely. There was something **unpleasant** about it, although she wasn't quite sure what. (BoE: BB_hF022228)
>
> She was paralysed by the scene itself, but, more than that, there was something **wrong** about her mother's movements, something dehumanized and robotic. (BoE: BB_aM941193)

As the three corpus instances show, it is possible to claim that the evaluative meaning of the target pattern is determined by the adjectives following *something* (i.e. *odd, unpleasant* and *wrong*). All the three adjectives share an evaluative meaning which describes something abnormal or not good. However, Hunston and Francis (2000) argued that the evaluative meaning should be granted to the pattern instead of the adjectives. For instance,

> Whether you're swanning about at Glyndebourne or sitting in a poppy field, there's something **very British** about picnics, especially if you use a traditional wicker basket. (Hunston and Francis, 2000, p.105)

Unlike adjectives such as *odd, unpleasant* and *wrong*, the adjective *very British* has a potentially neutral meaning, but the pattern still has an evaluative function. This indicates that it is not the words with which it occurs but the pattern itself that carries an evaluative meaning (Hunston and Francis, 2000).

The two arguments from Hunston and Francis's (2000) study indicate the interdependence between patterns and lexis, as well as the association between pattern and meaning. Hunston (2007, p.59) therefore summarizes that a pattern can either be

"a particular lexical item having a particular pattern available to it" (e.g. *afraid* frequently appears in the **ADJ *of*** and **ADJ *that*** patterns), or "a particular pattern selecting a given set of lexical items, or items with a particular meaning" (e.g. the verb slot in the pattern **V *over* n** can be filled by verbs in the 'argue' group). As a result, patterns should be treated as phraseological units because they integrate single words into grammatical structures, and the lexis and grammatical patterns come together to form units of meanings (Römer, 2009). As a phraseological unit, pattern grammar suggests that lexis and grammar are not isolate properties of the language (Francis, 1993; Hunston and Francis, 2000; Hunston, 2007).

Previous studies focusing on pattern grammar tended to investigate the patterns associated with individual verbs, nouns, and adjectives (e.g. Francis, 1993; Francis et al., 1996; 1998) or the meaning of a particular pattern (e.g. Groom, 2005; Su, 2019). However, the patterning of lexicalised multi-word combinations, such as PVs, has long been neglected. As lexicalised units, PVs have some similar functions with individual verbs. More specifically, PVs can form the verb clause in a sentence, and they are usually polysemous (Section 2.1.3). Based on these similarities between PVs and individual verbs, it would be possible to make two claims:

(1) if different meanings of a word can be distinguished by the patterns in which the word appears (e.g. different meanings of *reflect* appear in different patterns), it seems reasonable to assume that different meanings of a PV can also be distinguished by the grammatical patterns in which the PV appears.

(2) since words from the same meaning group are likely to appear in the same pattern (e.g. verbs in the 'argue' group are able to fill the pattern **V *over* n**), it seems that PVs, which share a similar aspect of meaning with one-word verbs,

are possible to be used in the same pattern in which the one-word verb equivalent appears.

The current study aims to fill the gap by investigating the grammatical patterning of PVs. Chapter 7 will present the findings about the patterns of PVs in the BoE corpus.

3.2.3 Identifying grammatical patterns in the corpus

Since the current study attempts to identify grammatical patterns associated with specific PVs, it is necessary to determine what criteria can be used to identify patterns in the corpus. The first criterion involved in identifying patterns is frequency (Hunston and Francis, 2000; Hunston, 2014). As mentioned in Section 3.2.1, the nature of pattern is repetition. If a group of words co-occurs with a particular grammatical structure repeatedly, then researchers could record the co-occurrence as a pattern (Hunston and Francis, 2000). The repeated use of a pattern can be manifested by its high frequency (at least twice) in the corpus (Hunston, 2010). However, Hunston and Francis (2000) noted that the frequent co-occurrence of words and a particular grammatical structure did not always indicate the existence of a pattern. Considering the following two corpus examples illustrated by Hunston and Francis (2000, p.69).

…which grow up to 15 cm tall and <u>are big enough to catch small fishes</u>.

It <u>is easy enough to make</u> if you buy the pastry…

As the two examples show, it is frequent for the word *enough* to follow an adjective and precede a to-infinitive clause in the corpus. However, Hunston and Francis (2000) questioned whether the two examples demonstrated the presence of a similar pattern. The authors argued that the pattern identified in the first example (i.e. **adj *enough* to-inf**) was unique to *enough*, whereas the pattern identified in the second should be ***it*** **v-link ADJ to-inf**, which was a pattern associated with adjectives (Hunston and Francis,

2000). This was because the to-infinitive clause was attribute to the adjective *easy* in the second example (Hunston and Francis, 2000). More specifically, if the word *enough* is missing in the second example, the absence does not affect the pattern and the meaning of the sentence. As a result, while corpus tools help to recognise frequent co-occurrence of words and grammatical structures, manual checks are still necessary because relying on frequency alone may generate inaccurate results (Hunston and Francis, 2000).

The second criterion is dependency, which means that "the pattern is not there by accident but as a result of the choice of the core word" (Hunston, 2014, p.102). Hunston (2014) used the verb form *recovering* to illustrate how dependency worked to identify patterns. In the corpus, this verb form is frequently followed by two prepositions, *from* and *in*. Hunston (2014) argued that *recovering from* can be identified as a pattern because the occurrence of the prepositional phrase beginning with *from* was dependent on *recovering*. The following two examples demonstrate the mutual dependence of *recovering* and *from*.

> Anyway, the Maiden are now ***recovering from*** the sad loss of Brucey baby.
>
> Frederick, at present ***recovering from*** an operation for throat. (Hunston, 2014, p.102)

However, this kind of connection cannot be found when *recovering* is followed by prepositional phrase beginning with *in.* For instance,

> He is *recovering in* bed.
>
> My nephew is *recovering in* hospital from multiple… (Hunston, 2014, p.102)

Hunston (2014) explained that the loose connection between *recovering* and *in hospital* can be tested by rephrasing.

In hospital, my nephew is ***recovering*** from multiple…

However, it is not acceptable to change the order of the sentences when *recovering* is followed by *from*.

From* an operation for throat, Frederick, at present ***recovering*.

As a result, when identifying a particular pattern with a word in a large collection of texts, the researcher needs to account for both frequent occurrences of the pattern and the dependency between the pattern and the node word (Hunston and Francis, 2000; Hunston, 2014). While corpus tools can present the frequency results of a grammatical structure, the question of whether the frequent grammatical structure is dependent with the node word seems to be difficult to answer. While rephrasing could be a useful test, it seems that the test still heavily relying on researchers' intuition (Hunston and Francis, 2000).

3.3 How the progressive form affects the meaning of verbs

The final section in this chapter is concerned with tense forms of verbs. It seems not feasible to capture the full range of meanings and functions conveyed by different tenses and aspects in English within a single section of a thesis. For the sake of space, this section cannot provide an exhaustive overview of all tenses and aspects in English but focuses on some representative studies regarding the progressive form of verbs.

At first glance, the connection between verb tense forms and the lexico-grammatical patterning of words appears to be weak. However, they are possible to be integrated if we examine the co-occurrence of a verb with a specific tense, such as the present progressive tense. When I investigated the use of PVs in the corpus, there was an interesting finding that some PVs tended to be frequently used in the progressive form

(e.g. *is looking up*). The frequent co-occurrence of specific PVs with the progressive form can be seen as a typical example of patterning. I will return to this point with more corpus examples in Chapter 6 (Section 6.2). Here in this section, it would be necessary to understand the semantic and pragmatic functions of using the progressive form.

Before moving on to the discussion of the functions of the progressive form, it is important to address potential confusion about the use of terms related to verb tense forms. In English grammar, there are two frequently used terms, 'tense' and 'aspect'. Biber et al. (1999) describe that while 'tense' refers primarily to the time orientation (i.e. past, present, and future), 'aspect' is associated with "considerations such as the completion or lack of completion of events or states described by a verb" (p.460). The progressive, which is also known as the continuous, is one aspect of verbs in English. However, corpus linguists may not use these two terms frequently. In corpus linguistics, the frequently used term seems to be 'form' or 'word form', which refers to different members of a lemma[16], including inflections for tense, number, case, etc. Since I focus on the *-ing* form of verbs in the corpus, the current study adoptes the label 'the progressive form'. Where 'the progressive form' is used, it refers to the structure the progressive aspect, which consists of the progressive auxiliary *be* followed by the present participle of a verb (e.g. *is looking*, *was looking*, and *have been looking*)

There are two aspects in English, the progressive and the perfect. Quirk et al. (1985) explained that the basic difference between the two aspects was to consider whether the action should be viewed as complete (i.e. perfective) or incomplete (i.e. progressive).

[16] According to Sinclair (1991), the lemma refers to all word forms of a word. For example, the lemma GO has different word forms including *go, goes, going, went, gone* and *to go*.

According to Biber et al.'s (1999) definition, the progressive aspect describes "activities or events that are in progress at a particular time, usually for a limited duration" (p.470). This definition of the progressive aspect agrees with Quirk et al. (1985), who also highlight the nature of the progressive aspect is "in progress" (p.189). Moreover, Biber et al. (1999) described that the present progressive aspect differed from the past progressive in the indicated time orientation. The present progressive aspect describes the events that are "currently in progress or about to take place in the near future", whereas the past progressive aspect describes events that "were in progress or about to take place at some earlier time" (Biber et al., 1999, p.470). Quirk et al. (1985) mentioned that there was an additional type, the prefect progressive, which integrated the two aspect constructions (i.e. *have been doing* or *had been doing*). However, in Biber et al.'s work, they did not account for the perfect progressive because this form was quite rare in the Longman Spoken and Written English Corpus (Römer, 2005).

In terms of the relationship between the progressive form and the meaning of verbs, Crystal (2004) noted that it may add semantic nuances to the meaning of verbs. Quirk et al. (1985) also investigated the association between the progressive form and three types of verb meanings. With the first type of verb meaning, Quirk et al. (1985) explained that stative verbs, such as *own, live,* etc., were not acceptable with the progressive aspect in most cases. While it is still possible to use the progressive form of stative verbs, some semantic nuances seem to be noticeable when this form is used (Quirk et al., 1985; Crystal, 2004). For example,

> We ***live*** in the country. vs. We ***are living*** in the country. (Quirk et al., 1985, p.199)

Based on the above examples, Quirk et al. (1985) argued that where the progressive

form of stative verbs was used, the meaning implied temporariness instead of permanence. The second meaning type is concerned with duration. More specifically, the progressive form indicates that the event "has duration and has not yet come to an end" (Quirk et al., 1985, p.199). The 'duration' meaning, which is also known as the 'incompletion' meaning (Mindt, 2000), can be illustrated by the example "[the] referee *is blowing* his whistle" (Quirk et al., 1985, p.199). Thirdly, in the example "[the] professor *is typing* his own letters while his secretary is ill" (p.199), the progressive form indicates a habitual meaning, implying a repeated action (i.e. *typing letters*) over a limited period of time (Quirk et al., 1985).

Quirk et al.'s work shows the three major meanings indicated by the progressive form (i.e. temporariness, duration and habit). Mindt (2000), who adopted a corpus-based approach to English grammar, found that the three meanings of the progressive from accounted for the majority of the distribution in the corpus. In addition to the three frequent meanings, Mindt (2000) observed that six additional meanings of the progressive form, such as "highlighting/prominence, emotion, politeness, prediction, intention, and matter-of-course" (see also Mindt, 1997, p.231). These additional meanings, although less frequent, appear to reflect the speakers' communicative purposes.

Some of the less frequent meanings, such as the 'emotion' meaning, were also mentioned by Quirk et al. (1985). Quirk et al. (1985) explained that the progressive form may lose the temporariness meaning when it was combined with adverbs such as *always, continually*, etc. For example,

> Bill *is* ***always*** *working* late at the office. (p.199)

With the use of *always*, Quirk et al. (1985) stated that the progressive form of *working* tended to describe an emotional feeling of disapproval. This finding is in line with Mindt's findings on the 'emotion' meaning of the progressive. However, besides the function of expressing negative emotions (e.g. disapproval and irritation), Mindt (1997) argued that the progressive form was also able to express positive emotions, such as agreement or sympathy. For instance,

I*'m longing* to meet my son-in-law. (p.231).

In this example, the progressive form maintained its emotional meaning without the explicit use of *always*. This is probably due to the fact that different meanings of the progressive frequently overlap, often encompassing two or more meanings simultaneously (Mindt, 1997).

When more than one meaning of the progressive form co-occurred in an utterance, Quirk et al. (1985) suggested that it sometimes conveyed tentativeness. This tentative meaning can be illustrated by the example "I *am hoping* you will come" (Quirk et al., 1985, p.202). As a stative verb, when *hope* was used in its progressive form, it typically implied temporariness, suggesting that *hoping* was a temporary state (i.e. the first major meaning of the progressive aspect). However, Quirk et al. (1985) noted that the use of the progressive form in this context added a sense of tentativeness to the utterance, making it sound more polite. This special use was labelled as the 'politeness' meaning of the progressive in Mindt's work (cf. Mindt, 1997; 2000). Mindt (1997) explained that using the progressive form can "soften the effect on the addressee", making the utterance less direct and harsh.

In addition to Mindt (1997, 2000) who focused on the communicative functions of the

progressive form, Römer (2005) also used a corpus-driven approach to investigate the various functions of the progressive form in spoken English. This study found that experienced speakers used the progressive form to "convey their (mostly negative) attitude to something" (p.99) especially when it co-occurred with *always* (Römer, 2005). This finding suggests that the emotional meaning of the progressive form is frequently negative but could be neutral or positive in some cases. Unfortunately, Römer (2005) did not show the neutral or positive emotion with corpus evidence. In a more recent corpus-based study on the progressive form, Lindley (2020) focused on the co-occurrence of *always* and the progressive form in two corpora, COCA and MICASE (the Michigan Corpus of Academic Spoken English). Lindley (2020) provided empirical data to argue that the emotionally neutral function (typically describing facts) was the most frequent function of the progressive form, accounting for 71.9% of occurrence in the corpora. Besides the neutral function, the negative function (typically complaints and laments) was more frequent than the positive function (typically praise) (Lindley, 2020). In the two corpora, 24.1% of the occurrences demonstrated the negative function whereas only 4% showed the positive function (Lindley, 2020).

Based on the review of previous studies on the progressive form, the fundamental functions of this form are related to the indication of a temporary, durative, or habitual state (Quirk et al., 1985; Mindt, 1997; 2000). In addition to the three major functions, this form is also able to reflect speakers' emotion and tentativeness (Mindt, 1997; 2000; Römer, 2005). As a result, the current study will examine the co-occurrence of the progressive form and specific PVs, analysing whether the use of progressive form have any semantic and pragmatic meanings with certain PVs.

3.4 Summary

This chapter has outlined the phraseological aspects that are involved in the current study. Previous studies tend to focus on the multiple meanings and the flexible syntactic structure of PVs. As a lexical item, there is little research on the phraseological use of PVs. This study tries to fill the gap by exploring the lexical-grammatical patterns associated with PVs. Based on the overview of collocation and pattern grammar in this chapter, the study will investigate nominal collocates and grammatical patterns occurring around PVs in the BoE corpus. It also examines whether lexico-grammatical patterns associated with PVs help to differentiate the meanings of PVs. In the meantime, it also looks at the progressive form of specific PVs (see Section 6.2), examining the association between the meanings of PVs and this special tense form.

Chapter 4 Methodology

This chapter aims to describe the methodological procedure that was employed in the current research. On the basis of the rationale explained in Section 1.2, Section 4.1 outlines the three research questions of the thesis and describes how each research question was addressed in the following result chapters. Then, Section 4.2 introduces the selection of corpora for this study. There are two reference corpora involved: the BoE corpus and a specialised sub-corpus extracted from the BNC. More details about how I use these two reference corpora can be found in Section 4.2.1. The Chinese learner corpus is derived from two existing learner corpora of Chinese EFL learners, with a careful selection of texts from both. Section 4.2.2 provides more information about the Chinese learner corpus. The following section (Section 4.3) shows the procedures of collecting data from the corpus. Finally, the analytical methods of the current study (i.e. quantitative and qualitative analysis) are presented in Section 4.4. This section also discusses the difference between the two types of analysis in detail.

4.1 Research Questions

As mentioned in Section 1.2, the main objective of this study is to explore how the use of PVs differs between Chinese EFL learners' writing and L1 users' writing. To achieve this objective, I believe it is necessary to first understand how PVs are used by L1 users and then compare the use of PVs in the Chinese learner corpus with what I found in the reference corpus. Therefore, the current study aims to answer three major research questions in the following chapters (Chapter 6 to Chapter 8):

1. How do L1 users use specific PVs in terms of three aspects: collocation, patterns, and the progressive form?

2. How does the use of frequent PVs (both two-word and three-word PVs) differ between the reference corpus and the Chinese learner corpus?
3. Do Chinese EFL learners replace PVs with one-word equivalents in their writing? When Chinese EFL learners replace PVs with one-word verbs, how satisfactory are their replacements?

As mentioned in Section 3.1 and 3.2, previous studies on collocations and grammatical patterns often investigated the collocational and grammatical patterns of individual words, whereas the patterns that co-occur frequently with PVs seem to be an area that is under-investigated. As a result, the first research question aims to fill this gap by investigating the lexico-grammatical patterns associated with specific PVs in the BoE corpus and exploring how different meanings of the PVs are associated with their frequent collocates and grammatical patterns. In addition to collocations and grammatical patterns, it seems that some PVs are frequently used in a particular form (e.g. the *-ing* form). I am also interested in whether the use of specific forms influences the meaning of the PVs. Answers to this research question attempt to provide insights into how L1 users use PVs and to enhance comprehension of polysemous PVs with respect to the lexico-grammatical patterns associated with these PVs.

The second research question aims to examine whether Chinese EFL learners under- or overuse frequent PVs in their writing in comparison to L1 users. To answer this question, I focused on a list of 100 frequent two-word PVs proposed by Gardner and Davies (2007) and investigated the frequency and distribution of these PVs in the reference corpus and the Chinese learner corpus. Besides overall frequency, the results of this question also showed the distribution of different forms (i.e. the contiguous and non-

contiguous form) of these 100 PVs in the two corpora, examining whether the distribution of each form in the Chinese learner corpus differs from the distribution in the reference corpus. When answering this question, I also paid attention to the use of specific PVs in the Chinese learner corpus, examining whether Chinese EFL learners are able to use these PVs in a semantically and syntactically appropriate way. Based on this qualitative analysis, it is possible to develop an understanding of Chinese EFL learners' difficulties of using PVs

Finally, the third research question investigates Chinese EFL learners' choices between PVs and synonymous one-word verbs. As mentioned in Section 2.3.1, previous studies have argued that many EFL learners tend to avoid using multiword verbs and prefer to replace them with one-word verbs (Siyanova and Schmitt, 2007; Waibel, 2007). Based on this argument, this research question investigates whether Chinese EFL learners also have such avoidance behaviour on their use of PVs. In order to limit the scope of the study, the study selected 10 PVs and their one-word verb equivalents and examined the use of these 10 verb pairs in the Chinese learner corpus.

4.2 Corpora and corpus tools

In order to answer the three research questions, this study employs a corpus-based method to investigate how experienced users of the language use PVs (using data taken from the reference corpus) and to compare how Chinese learners' use of PVs differs from L1 users (using data taken from both the reference corpus and the learner corpus).

Before introducing the corpora used for investigation and comparison, it is important

to note that large corpora are desirable for the current study. In order to establish the lexico-grammatical patterning associated with the PVs (i.e. collocation and grammatical patterns), it is necessary to use a large corpus, especially a large learner corpus. Having access to a large learner corpus is beneficial for retrieving longer sequences of words in the corpus. Sinclair (2004b, p.189) argued that "the main virtue of being large in a corpus is that the underlying regularities have a better chance of showing through the superficial variations". O'Keeffe and MaCarthy (2010) agreed with this idea and acknowledged that large corpora were suitable for the investigation of lexis and phraseology. As a result, when investigating the use of specific PVs in longer sequences, a learner corpus with a considerable size would be beneficial as it increases the opportunities of finding repeated patterns associated with PVs.

Another reason for finding large corpus is associated with the stylistic feature of English PVs. As mentioned in Section 2.2.3, PVs are more common in spoken language than in written language. Therefore, it can be assumed that the overall frequency of PVs in written texts may not be as high as that in spoken texts. A larger collection of written texts would be beneficial for data collection, providing more corpus evidence of learners' use of PVs.

4.2.1 The reference corpora

The current study involves two reference corpora, the Bank of English (BoE) corpus[17] and a specialised sub-corpus derived from the British National Corpus (BNC). The BoE corpus was used for investigating how experienced users use PVs, as it is one of the

[17] More information about the Bank of English corpus is available at the website: https://cqpweb.bham.ac.uk/bankofenglish/

largest reference corpora designed to reflect the mainstream of contemporary English (Mahlberg, 2005). To be more specific, the BoE corpus is mainly associated with the first research question and provides corpus data about the lexico-grammatical patterning associated with PVs in English. In terms of research questions 2 and 3, a more specialised corpus of written English was used in order to compare the use of PVs in L1 users' writing with that in Chinese learners' writing.

The BoE corpus, containing both written and spoken texts of English, is a large-scale reference corpus which is "regularly updated and has been continually increasing in size" (Mahlberg, 2005, p.42). When I started investigating the use of PVs in the BoE corpus, its size had reached over 700 million running words[18]. The large size of the BoE corpus indicates that it is a representative corpus, providing a comprehensive view of how L1 users use PVs in both written and spoken discourse. In addition to size, the 'representativeness' of this corpus is also associated with geographic diversity (Hunston, 2002; Mahlberg, 2005). Hunston (2002) describes that the idea of geographic diversity refers to the equal collection of data from various regions. Table 4.1 below provides information about the sub-corpora of the BoE corpus (cf. Barnbrook, Mason, and Krishnamurthy, 2013, p.214). As Table 4.1 shows, although the BoE corpus predominantly collects texts from British English, it also includes texts from other regions such as the United States, Australia, Canada, etc. (Barnbrook, Mason, and Krishnamurthy, 2013). The wide selection of texts from different regions also ensures that searching for PVs in the BoE corpus would yield more comprehensive results. The BoE corpus is also intended to be a resource where dictionary writers and learners of

[18] The BoE corpus was last updated in 2012 on CQPweb, the web-based corpus tool where I accessed the corpus.

English could expect to find standard English (Hunston, 2002). As a result, the current study considered the BoE corpus a valuable resource for consultation, providing instances of how experienced users use PVs.

Table 4.1 Sub-corpora of the BoE corpus (cf. Barnbrook, Mason, and Krishnamurthy, 2013, p.214)

Sub-corpora	Country of origin	Contents
oznews	Australia	Australian newspaper
strathy	Canada	Canadian mixed corpus
bbc	UK	BBC radio
brbooks	UK	British general books
brephem	UK	British ephemera
brmags	UK	British magazines
brspok	UK	British informal spoken language
econ	UK	*Economist* magazine
guard	UK	*Guardian* newspaper
indy	UK	*Independent* newspaper
newsci	UK	*New Scientist* magazine
sunnow	UK	*Sun/News of the World* newspaper
times	UK	*Times/Sunday Times* newspaper
wbe	UK	Business language
npr	US	National Public Radio
usacad	US	American academic books
usbooks	US	American general books
usephem	US	American ephemera
usnews	US	American newspapers
usspok	US	American information spoken language

Despite the usefulness of the BoE corpus, this study did not use it for comparison with the learner corpus. I was looking for another reference corpus for comparison because the size of the BoE corpus is too large to find a suitable learner corpus. It was not easy to find a learner corpus of comparable size. In addition, as Table 4.1 shows, the BoE corpus consists of around 90% written texts and 10% spoken texts. It could be argued that spoken texts, such as newspaper articles, magazine articles, radio podcasts (see Table 4.1), are still largely scripted, making the spoken sub-corpus not entirely

reflective of oral language. Furthermore, the BoE corpus is sometimes criticised for an unbalanced inclusion of texts, relying heavily on journalistic texts (Mahlberg, 2005). Although journalistic texts are able to represent mainstream English, the genre differences between journalistic texts and learners' English writing suggest that the current study should not treat the BoE corpus as the reference corpus for comparison.

A more trustworthy comparison can be made using the BNC, another general-purpose corpus of English. The full size of the BNC contains approximately 100 million words, with 90% of the data coming from the written component and 10% from the spoken component (Burnard, 2007). In contrast to the BoE corpus which relies largely on journalistic texts, the written sub-corpus of the BNC has a balanced selection of texts across various genres, such as academic and non-academic prose, fictions, newspaper articles, etc. (Hunston, 2002). However, it is still too large to be a reference corpus of comparable size. As a result, considering the size of the Chinese learner corpus (see Section 4.2.2 for more information), I decided to select three text files from the BNC and make up a reference corpus of comparable size based on the selected files. Details of the reference corpus used for comparison can be found in Table 4.2.

Table 4.2 Text files included in the reference corpus for comparison (selected from the BNC

Text files	Mode	Place of origin	Tokens	Texts
Non-academic proses			3,974,099	117
School & University essays	Written	UK	188,335	9
Letters			111,178	8
Total			4,273,612	134

The reference corpus used for comparison contains 4,273,612 tokens (running words) in 134 texts originated from the UK. While the reference corpus for comparison is much

smaller than the full version of the BNC, it has a comparable size with the learner corpus (2,882,429 words). Meanwhile, to avoid invalid comparison between the reference corpus and the learner corpus, I made a careful selection of text files. Some text files in the BNC, such as academic prose, fiction, and newspapers, were excluded from this specialised corpus. This exclusion was due to the absence of these genres in the learner corpus (see Table 4.3 in Section 4.2.2). The selected text files in the comparable reference corpus aim to be consistent with those included in the Chinese learner corpus.

In this thesis, corpus evidence from the BoE corpus mostly appeared in Chapter 6 and Chapter 7 where the BoE corpus can demonstrate the lexico-grammatical patterns associated with PVs. Chapter 2 and Chapter 5 also included some examples from the BoE corpus when discussing the identification of English PVs in the corpus. In terms of the comparison between L1 users and EFL learners, the BoE corpus was discarded. The specialised sub-corpus derived from the BNC provided data to show how the 100 most frequent PVs were used in L1 users' writing. For the purpose of simplicity, the reference corpus in this thesis particularly refers to the specialised sub-corpus derived from the BNC rather than the BoE corpus, because I did not compare the use of PVs in the BoE corpus and the Chinese learner corpus.

To access and investigate data in the BoE corpus and the reference corpus, I used CQPweb[19] (Hardie, 2012), a web-based software for searching a corpus and analysing corpus data. This is a powerful corpus analytical tool with a variety of useful functions. For example, Figure 4.1 is a screenshot showing the query results of *LOOK up*[20] in the

[19] More information about CQPweb can be found at the website: https://cqpweb.bham.ac.uk/index.php.
[20] Small capital letters are used to refer to all members of a lemma. *LOOK up*, for example, refers to *look*

BoE corpus. As Figure 4.1 shows, CQPweb allows users to analyse this PV by looking at the collocations, the frequency of different forms, the distribution of it in the corpus, etc.

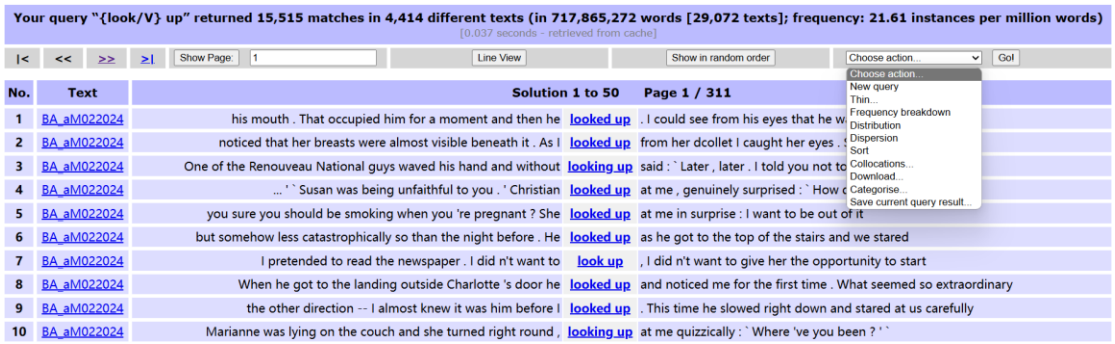


Figure 4.1 A screenshot of CQPweb showing the functions of this corpus tool

4.2.2 The Chinese learner corpus

The learner corpus in the current study is a combined corpus of two ready-made learner corpora for Chinese EFL learners: the WECCL (Written English Corpus of Chinese Learners) (Wen, Liang, and Yan, 2008) and the TECCL (Ten-thousand English Compositions of Chinese Learners) (Xu, 2015).

The WECCL corpus is a sub-corpus of the Spoken and Written English Corpus of Chinese Learners (SWECCL 2.0) (Wen, Liang, and Yan, 2008), one of the most significant learner corpora for research on Chinese EFL learners (Zou and Reinders, 2017). The SWECCL corpus contains about 2 million words and consists of two sub-corpora: the spoken sub-corpus and the written sub-corpus. Since the current study focuses on the use of PVs in written texts rather than spoken texts, I used data from the written sub-corpus of SWECCL and excluded data from the spoken sub-corpus from my learner corpus. When establishing the written sub-corpus, Wen, Liang, and Yan

up, looked up, looking up, and *looks up*. In this thesis the lemma is used with the lexical verbs in PVs rather than the particles.

(2008) aimed to create a corpus that more accurately and objectively reflects the English writing proficiency of Chinese EFL learners. The WECCL corpus is a collection of 4950 written texts from Chinese EFL learners, containing 1,186,215 words. The texts were written by college students from over 20 universities in mainland China, including both English majors and non-English majors (Wen, Liang, and Yan, 2008). The texts covered 27 topics, consisting of 26 argumentative essays and one expository essay (see Appendix 3 for more details about these topics). In addition, essays in the WECCL were written under time constraints. The majority of the texts were exam essays written for the TEM-4 test. TEM-4 is targeted at first- and second-year undergraduates with an intermediate level of English proficiency (Jin and Fan, 2011). Although the essays were written under exam conditions, they were not targeted at academic audience, such as professors or tutors of a specific discipline. Students were asked to discuss a topic they may be confronted with every day and make some arguments based on their personal experience (Guo and Chen, 2007). Based on Guo and Chen's (2007) description, it could be argued that the written texts in the WECCL corpus were non-academic writing and did not exhibit features of academic writing.

The TECCL corpus is another learner corpus involved in the current study. Similar to the WECCL corpus, the TECCL corpus is also a corpus of students' writing produced by Chinese EFL learners. Created in 2015, the TECCL corpus has been more recently updated compared with the WECCL corpus. The size of the TECCL corpus is over 1.8 million words (1,817,472 words), which is also larger than the WECCL corpus. This corpus contains approximately 10,000 written texts from Chinese students, ranging from elementary school students to postgraduates (Xu, 2015). While the texts in this corpus were written by students with varying levels of English proficiency,

undergraduates with the intermediate proficiency level were the major contributors (Xu, 2015). Although text contributors to the TECCL corpus did not provide a full list of essay topics, Xu (2015) offered some prompts regarding the essays, which help to indicate the genres of learners' writing. Based on Xu's prompts, Table 4.3 displays some examples of essay topics in the TECCL corpus and shows the corresponding genres.

Table 4.3 Examples of essay topics in the TECCL corpus

Genres	Essay topics
Argumentative essays	- How to Build Up Self-confidence? - Discuss the positive and negative aspects of sports. - Should college students take part in social practice?
Expository essays	- Describe how the number of postgraduates increases over the years based on the table. - Describe the relationship between one's educational background and earnings based on the bar chart. - Describe the layout of the room based on the map.
Empathy texts (emails and letters)	- A letter inviting a hypothetical pen pal to visit Beijing during the Olympic Games. - An email checking the details for the seminar.

As shown in Table 4.3, I used the label 'empathy texts' to refer to emails and letters in the TECCL corpus. This label was borrowed from Nesi et al. (2012), who categorises job applications, letters to friends, news reports, etc. as 'empathy writing'. The authors also explain that producers of empathy writing "do not usually have to construct an argument or demonstrate new reasoning, but they do need to engage with their hypothetical readers or listeners" (Nesi et al., 2012, p.217). According to their explanation, 'empathy texts' should be a suitable genre for learners' emails and letters in the TECCL corpus, although this corpus did not include other types such as job application and news report.

The TECCL corpus collects texts written under various conditions, including exam essays, as well as compositions written after class and collaborative writing tasks. Also, the TECCL corpus includes essays written by English majors as well as non-English majors who are presumed to have less exposure to English than their English-major counterparts. Unlike the WECCL corpus, which focuses primarily on English-major undergraduates in China, Wu and Tissari (2021) believe that the TECCL corpus is more likely to represent the overall English proficiency of Chinese EFL learners. This is attributed to its extensive inclusion of student writing, which provides a more comprehensive view of English proficiency levels among Chinese EFL learners (Wu and Tissari, 2021).

Table 4.4 Information about two ready-made learner corpora targeted at Chinese EFL learners

	WECCL (Wen, Liang, and Yan, 2008)	**TECCL** (Xu, 2015)
Size	1,186,215	1,817,472
Year of collection	2003 – 2007	2011 – 2015
Number of texts	4950	9864 (with 6867 written by undergraduates)
Average length of the texts	150 - 250 words	150 – 200 words
Learners' L1	Chinese	Chinese
Learners' proficiency level	Intermediate	Beginners to upper-intermediate (with the intermediate being the majority)
Time conditions	Timed writing	Both timed and untimed writing
Types of learners' English writing	Argumentative essays and one expository essay	Argumentative essays, expository essays, emails and letters

Based on the description above, Table 4.4 summarises the information about the two learner corpora. As Table 4.4 shows, the two corpora have some common features. For

example, both of them collected non-academic writing from Chinese EFL learners with the same L1 background (i.e. Chinese). They also have some distinctions. Compared to the WECCL corpus, which focuses particularly on undergraduate students' exam writing, the TECCL corpus has a wider selection of texts, collecting essays written under time constraints and without time limit (see Table 4.4). Also, text contributors of the TECCL corpus have various levels of proficiency, ranging from beginners to upper-intermediate learners.

As mentioned at the beginning of Section 4.2, the current study requires a learner corpus that has a considerable size, allowing for the investigation of the lexico-grammatical patterns associated with PVs. To obtain a desirable large-size learner corpus, I decided to combine the two learner corpora and called the combined corpus the Chinese learner corpus (or the learner corpus) in the thesis. Considering the differences between the two learner corpora, I made a selection of students' writing in the two corpora before combining the two. Firstly, around 3000 out of the 9864 written texts in the TECCL corpus were excluded in the Chinese learner corpus. This is because they were not written by undergraduate students. The remaining 6867 texts in the TECCL corpus were produced by undergraduate students whose level of proficiency should be similar to the learners in the WECCL corpus.

Besides learners' proficiency level, I also considered the selection of essays under different conditions. Na and Yoon (2016) argue that time is a crucial factor that affects the quality of learners' writing. They found learners tend to perform better in the untimed condition (Na and Yoon, 2016). However, I decided to include both timed and untimed writing in the Chinese learner corpus because of two reasons: (1) in the TECCL

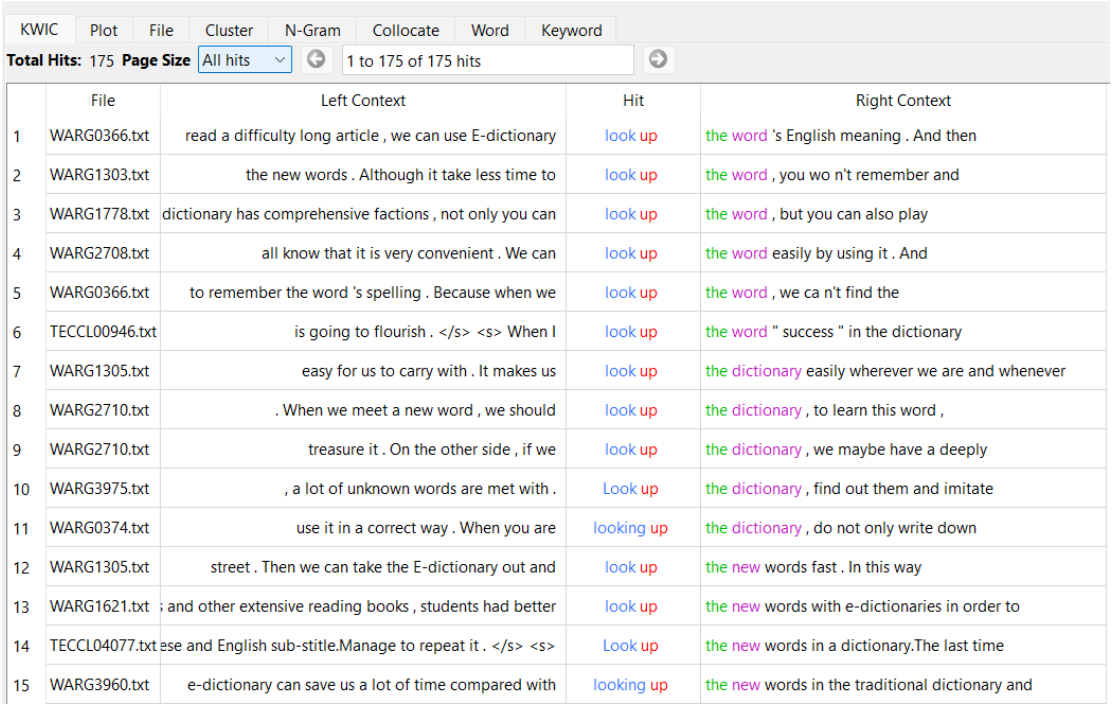
corpus, Xu (2015) did not specify which compositions were written under time constraints and which were not, and (2) the inclusion of both would be able to show learners' performance of using PVs under different conditions, yielding more comprehensive results.

Finally, in terms of text types, the Chinese learner corpus contains three different types of student writing: argumentative essays, expository essays, and empathy texts (emails and letters). Both the WECCL and the TECCL contain argumentative essays and expository essays (see Table 4.4), which are the most common text types of college student English writing in Chinese universities. In addition to the two types, the TECCL corpus also includes a small number of emails and letters written by undergraduates (see Table 4.3). Although empathy texts only exist in the TECCL corpus, I retained this type in the Chinese learner corpus because learners may employ informal language when they tried to engage with their hypothetical readers through emails or letters. In other words, the inclusion of emails and letters may provide better opportunities to analyse learners' use of PVs in written texts.

To sum up, the Chinese learner corpus in the current study contained 2,882,429 words. This size should be comparable with the reference corpus. Among 11,818 pieces of students' writing, both timed and untimed essays were included. Producers of the written texts were Chinese undergraduate students who speak Chinese as their first language. Learners' level of English proficiency was intermediate. The Chinese learner corpus contained learners' non-academic writing, consisting of argumentative essays, expository essays and empathy writing (emails and letters), which are consistent with the genres included in the reference corpus (i.e. non-academic prose, school and

university essays, and letters).

AntConc 4.2 (Anthony, 2023) was used in this study for analysing corpus data in the Chinese learner corpus. This was different from how I accessed the reference corpora via CQPweb. The reason for choosing a different corpus tool was that the Chinese learner corpus was unable to be loaded and compiled on CQPweb. As a consequence, I needed a corpus tool which could compile the corpus based on uploaded text files. AntConc is a multi-functional yet easy-to-use corpus analysis toolkit. In addition to compiling my learner corpus, this tool also has many functions that are useful for data analysis. For example, users can use the KWIC function to read the concordance lines of the queried item. Figure 4.2 is a screenshot showing the concordance results of *LOOK up* in the Chinese learner corpus. Besides showing KWIC concordance, users can also use other functions of AntConc, including analysing clusters, n-grams, or collocates, and generating word and keyword frequency, etc. Figure 4.3 is a screenshot showing the collocates of *LOOK up* in the Chinese learner corpus.

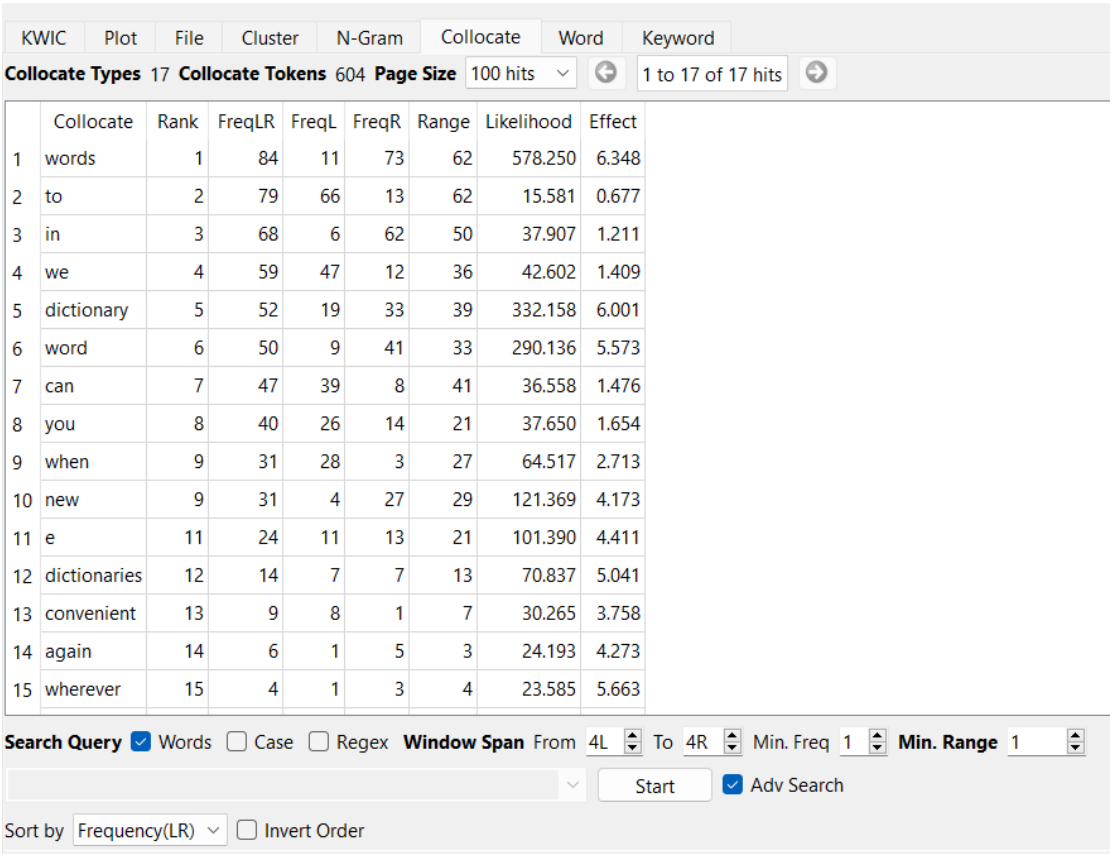

KWIC | Plot | File | Cluster | N-Gram | Collocate | Word | Keyword

Total Hits: 175 **Page Size** All hits | 1 to 175 of 175 hits

	File	Left Context	Hit	Right Context
1	WARG0366.txt	read a difficulty long article , we can use E-dictionary	look up	the word 's English meaning . And then
2	WARG1303.txt	the new words . Although it take less time to	look up	the word , you wo n't remember and
3	WARG1778.txt	dictionary has comprehensive factions , not only you can	look up	the word , but you can also play
4	WARG2708.txt	all know that it is very convenient . We can	look up	the word easily by using it . And
5	WARG0366.txt	to remember the word 's spelling . Because when we	look up	the word , we ca n't find the
6	TECCL00946.txt	is going to flourish . </s> <s> When I	look up	the word " success " in the dictionary
7	WARG1305.txt	easy for us to carry with . It makes us	look up	the dictionary easily wherever we are and whenever
8	WARG2710.txt	. When we meet a new word , we should	look up	the dictionary , to learn this word ,
9	WARG2710.txt	treasure it . On the other side , if we	look up	the dictionary , we maybe have a deeply
10	WARG3975.txt	, a lot of unknown words are met with .	Look up	the dictionary , find out them and imitate
11	WARG0374.txt	use it in a correct way . When you are	looking up	the dictionary , do not only write down
12	WARG1305.txt	street . Then we can take the E-dictionary out and	look up	the new words fast . In this way
13	WARG1621.txt	; and other extensive reading books , students had better	look up	the new words with e-dictionaries in order to
14	TECCL04077.txt	ese and English sub-stitle.Manage to repeat it . </s> <s>	Look up	the new words in a dictionary.The last time
15	WARG3960.txt	e-dictionary can save us a lot of time compared with	looking up	the new words in the traditional dictionary and

Figure 4.2 A screenshot of AntConc showing the concordance results of *LOOK up* in the

Chinese learner corpus.



	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	words	1	84	11	73	62	578.250	6.348
2	to	2	79	66	13	62	15.581	0.677
3	in	3	68	6	62	50	37.907	1.211
4	we	4	59	47	12	36	42.602	1.409
5	dictionary	5	52	19	33	39	332.158	6.001
6	word	6	50	9	41	33	290.136	5.573
7	can	7	47	39	8	41	36.558	1.476
8	you	8	40	26	14	21	37.650	1.654
9	when	9	31	28	3	27	64.517	2.713
10	new	9	31	4	27	29	121.369	4.173
11	e	11	24	11	13	21	101.390	4.411
12	dictionaries	12	14	7	7	13	70.837	5.041
13	convenient	13	9	8	1	7	30.265	3.758
14	again	14	6	1	5	3	24.193	4.273
15	wherever	15	4	1	3	4	23.585	5.663

Figure 4.3 A screenshot of AntConc showing the collocate function of this tool

4.3 Data collection procedure

4.3.1 Retrieving PVs in the corpus

As noted in Section 2.2, both two-word PVs and three-word PVs (i.e. phrasal-prepositional verbs) that function as a lexicalised unit will be investigated in the current study. Compared to retrieving three-word PVs, retrieving two-word PVs seems to be more challenging. This is because the structure of three-word PVs tends to be more stable than their two-word counterparts. Three-word PVs, such as *look forward to*, *keep up with*, etc., usually have a fixed structure, whereas two-word PVs, especially separable PVs (e.g. *take off*), have a more dynamic structure. Meanwhile, due to the fact that two-word PVs are more frequent in English, it is challenging to retrieve all two-word PVs in the corpus.

Considering the methodological challenge for retrieving two-word PVs, this study therefore investigates a list of 100 frequent PVs in the reference corpus and the Chinese learner corpus. These 100 PVs were not chosen randomly. This list was created by Gardner and Davis (2007) who studied frequent English PVs in the BNC (see Appendix 1 for the full list of these 100 PVs). These 100 frequent PVs had a high coverage in the BNC, accounting for more than half (50.4%) of all the occurrences of PVs (Gardner and Davis, 2007). Because of the high coverage, this list of high-frequency PVs is considered a useful resource for EFL learners (Garnier and Schmitt, 2015).

The first step was to prepare the list of highly frequent two-word PVs. Once the list was ready, I searched for the 100 PVs in the reference corpus and the Chinese learner corpus. As mentioned in Section 2.1, the basic structure of English two-word PVs is a lexical verb plus a prepositional or adverbial particle. The main task in the second stage was to retrieve concordance results that contained both the lemmatised verb and the particle in the two corpora. CQPweb and AntConc use different query language when retrieving lemmatised items. In CQPweb, for example, all members of the lemma *LOOK* can be retrieved by '{look}'. Since the current research was looking for the verb *LOOK* rather than other word classes, so the query in CQPweb should be refined to '{look/V}', which retrieved all verbal members of the lemma *LOOK*. In AntConc, however, the wildcard '||' helps to search for lemmatised items. Therefore, in order to retrieve the lemma *LOOK* in the Chinese learner corpus, the query should be 'look||looked||looking||looks'. Then non-verbal members of *LOOK* were manually eliminated from the results.

Another methodological challenge for retrieving two-word PVs in the corpus was

associated with their dynamic forms. As discussed in Section 2.1.3, some two-word PVs have a non-contiguous form. Gardner and Davis (2018) found that separable PVs typically had pronouns and noun phrases as the intervening elements between the verb and the particle. Due to this feature, the current study did not examine separable PVs that were separated by three or more intervening words. To be more specific, the threshold for the length of intervening items was two words. The reason for having this threshold was the low frequency of PVs when the distance between the verb and the particle exceeds two words. Figure 4.4 exhibits several concordance lines in which *LOOK* and *up* was separated by three intervening words. As Figure 4.4 shows, when three items appear between *LOOK* and *up*, it is less likely to identify phrasal use of *LOOK up* in the BoE corpus.

Moreover, the average person	**looks**	at their watch **up** to 50000 times a year
completely untroubled. I'll	**look**	forward to catching **up** with you after
Lower Refectory. She	**looked**	away and pulled **up** her scarf, trying to
though unemployed, are	**looking**	for work make **up** the total labour force
other direction he could	**look**	from his house **up** the street to the Museum

Figure 4.4 When three words intervene *LOOK* and *up* in the BoE corpus

In CQPweb, searching for non-contiguous word sequences requires the use of wildcards that define the distance between words in a query (Hardie, 2012). The non-contiguous form of *LOOK up* can be retrieved by adding the wildcard '+' and '*'. The wildcard '+' is used to skip an arbitrary token and '*' is for an optional token (Hardie, 2012). As Table 4.5 shows, the query '{look/V} + up' suggested that the verb and the particle were separated by one token, whereas '{look/V} ++ up' retrieved instances in which 2 tokens appeared between *LOOK* and *up*. In terms of '{look/V} +* up', this query retrieved instances in which *LOOK* and *up* were separated by one or two tokens. For example, both '*look* them *up*' and '*look* the number *up*' can be extracted using this query. It was helpful when I needed to retrieve all instances that contain the non-contiguous

form of *LOOK up*.

Table 4.5 Retrieving separable PVs (e.g. *LOOK up*) in CQPweb

Query	Concordance results
{look/V} up	She **looked up** at me in surprise. (BoE: BA_aM022024) We must keep in mind things are **looking up**. (BoE: NA4__040812)
{look/V} + up	They **looked** it **up** in a book. (BoE: BA_Qf022206) My first reaction was to **look** straight **up** to see whether the rig had survived. (BoE: BA_wm012092)
{look/V} ++ up	The PM **looks** the man **up** and down. (BoE: BA_SF042743) Trujillo **looked** the number **up** and dialled it. (BoE: BU_cM022233)
{look/V} +* up	**Look** them **up** on the internet. (BoE: NB1__040930) It is always possible to **look** these things **up** or make a phone call. (BoE: MBX____1161)

AntConc also enables users to search for a sequence of words, but this software relies on a different query system. When two intervening words separate *LOOK* and *up*, it also means that *up* appears on the right side of *LOOK* within a span of three words. Figure 4.5 is a screenshot demonstrating how I retrieved separable PVs in the Chinese learner corpus using AntConc.

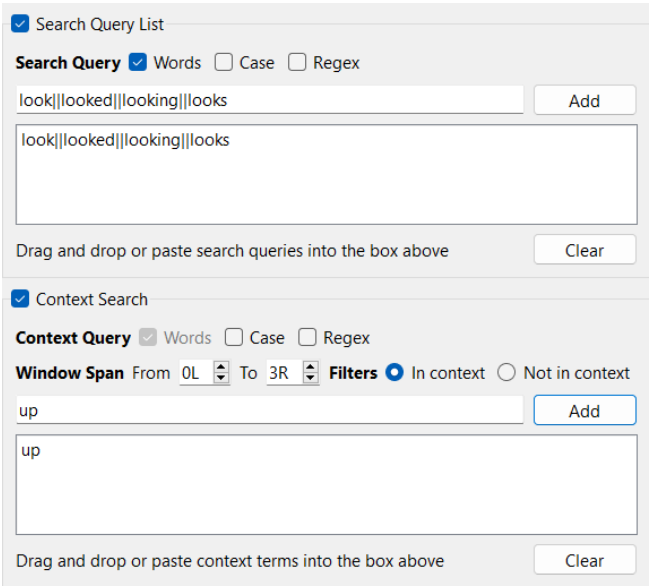


Figure 4.5 Retrieving separable PVs (e.g. *LOOK up*) in AntConc

4.3.2 Identifying PVs and rejecting non-PV combinations

When retrieving concordance results for the 100 frequent PVs, it is important to note that the data may not be sufficiently clean. Some free combinations of the verb and the particle might have been automatically retrieved. This is because the corpus tools could not guarantee that each concordance contained a verb + particle combination that is used as a PV. Figure 4.6 exhibits some concordance lines extracted from the reference corpus. In these concordance lines, the verb + particle combinations were not used as PVs.

```
                a nurse who  comes in the morning and at night
 Katherine's remains were  coming on the Saturday and on the Friday
          employers have  gone over the last ten or fifteen years
              I like to  put it on the table and hear the comments
insufficient planting has  taken place over the past thirty years.
```

Figure 4.6 Non-PV combinations in the reference corpus

In Figure 4.6, it is impossible to claim that the verbs and the particles were combined as units of meaning. For example, *comes in* cannot be regarded as a PV in the first line. This is because 'in the morning' is a prepositional clause that modifies the verb 'comes', specifying the time when the nurse comes. Consequently, after the first stage of retrieving verb + particle combinations in the reference corpus and the Chinese learner corpus, it was necessary to manually check each concordance result and eliminate those in which the verb and the particle did not form a unit of meaning.

When combinations such as 'comes in the morning' was eliminated, the remaining data should contain verb + particle combinations that were used as units of meaning. The next step was to determine whether these combinations should be classified as PVs or non-PV combinations. As noted in Section 2.2.1, using syntactic tests to identify PVs

tends to be unreliable because these syntactic tests are not applicable to inseparable PVs and intransitive PVs. Due to the weaknesses of syntactic tests, this study employed semantic criteria to identify PVs. Here is a brief summary of the two semantic criteria:

1. examining whether the meaning of a verb + particle combination is compositional or non-compositional (i.e. whether the meaning of the combination can be predicted by the meanings of the components);
2. examining whether one of the components, either the verb or the particle, has an extended meaning in the combination (i.e. whether the figurative or metaphorical meaning of the component helps to understand the meaning of the combination).

Since the focus of this section is to describe the steps of data collection, I am unable to show how specific PVs are identified using the two semantic criteria. In Section 5.1, I will provide more details and use corpus examples to demonstrate the identification of PVs. Compared with traditional syntactic tests that often ignore inseparable PVs and intransitive PVs, the two semantic criteria are able to identify both of them (see examples in Section 5.1). Also, the semantic criteria are applicable to both two-word and three-word PVs.

When I was dealing with the corpus data, an important observation was that verb + particle combinations usually have multiple usages. That is, a verb + particle combination can be used either as a PV or as a non-PV combination. Figure 4.7 below provides two random concordance results of *GO out* in the reference corpus. *GO out* was used in different ways in the two examples. In the context 'heating boiler which *goes out*', the combination was classified as a PV, whereas it was classified as a non-PV combination in 'I don't *go out* in the evenings'. Using the semantic criteria is able to

differentiate between PV use and non-PV use. I will return to the variation in the use of verb + particle combinations in Section 5.2.

```
heating boiler which goes out every time the money runs out
            I don't go out in the evenings, Mother doesn't like it
```

Figure 4.7 Two random concordance lines of *GO out* in the reference corpus

4.3.3 Finding synonymous one-word verbs of PVs

The final stage of data collection was to find one-word verbs whose meanings are similar to corresponding PVs. Considering the scope of the study, I selected 10 PVs and consulted WordNet about their one-word synonyms. WordNet[21] (Princeton University 2010) is a lexical database of semantic relations between words. It shows the semantic links between words including synonyms, hyponyms, and meronyms. The 10 PVs and their one-word equivalents are shown in Table 4.6.

Table 4.6 The selected 10 PVs and their one-word verbs

PVs	One-word verbs
GET back	return
GO up	rise
HOLD on	wait
LOOK for	search
PUT off	postpone
SET up	establish
TURN down	decrease
WORK out	solve
CATCH up with	reach
GET out of	avoid

The 10 PVs listed in Table 4.6 were selected for several reasons. Firstly, they were frequently used both in the reference corpus and the Chinese learner corpus. Out of the 10 PVs, 7 were taken from the list of 100 frequent PVs, such as *HOLD on* and *WORK out*. The rest three (i.e. *LOOK for*, *CATCH up with*, and *GET out of*), although not included in

[21] More information about WordNet can be found at the website: https://wordnet.princeton.edu/

the list, they were also frequently used in the two corpora. Secondly, the 10 PVs were intended to represent the range of PVs a learner may be exposed to. Both two-word and three-word PVs were included. They also varied in form and the compositionality of meaning. Some of them could be used separately (e.g. *PUT off, WORK out*), while some were more likely to be used in a contiguous form (e.g. *GO up, LOOK for*). Some of them had a compositional meaning (e.g. *GET back, GO up*), while others were not (e.g. *LOOK for, CATCH up with*). Thirdly, in the 10 selected PVs, some of them had multiple meanings (e.g. *SET up, TURN down*) while others did not (e.g. *HOLD on*). For each of these 10 PVs, a one-word alternative that is roughly synonymous was chosen (e.g. *GO up*/*rise*, *TURN down*/*decrease*).

It should be noted that these near synonyms may not be perfect alternatives to PVs, especially to PVs with multiple meanings. For example, the meaning of *TURN down* in '*turn down* the heat' is different from its meaning in '*turn down* his request'. The former is synonymous with 'decrease', whereas the latter seems to be associated with 'refuse'. As a result, this study has a limitation that the chosen one-word alternatives may not be perfect replacements for PVs and learners may have different choices if they intend to avoid using these polysemous PVs in their writings.

To sum up, in the current study, the procedure for data collocation involved three stages:

(1) retrieving concordance lines that include the 100 frequent PVs in the reference corpus and the Chinese learner corpus;
(2) cleaning the corpus data and using semantic criteria to classify verb + particle combinations as either PVs or non-PV combinations;
(3) finding 10 pairs of PVs and one-word equivalents and comparing the use of

these pairs in the two corpora.

In the first stage, I retrieved verb + particle combinations in the reference corpus and the Chinese learner corpus. However, the preliminary data may not be sufficiently clean because the corpus tools were unable to distinguish between PVs and non-PV combinations. The aim of the second stage was to manually clean the data and identify PVs from the concordance results. Since verb + particle combinations may have multiple usages in the two corpora, the current study classified verb + particle combinations as PVs or non-PV combinations using semantic criteria. After cleaning the corpus data, the remaining concordance results had been carefully recorded for further quantitative and qualitative analysis. Finally, in order to investigate whether Chinese EFL learners have a tendency to avoid using PVs, 10 PVs and their one-word equivalents were chosen. The concordance results of these 10 pairs in the two corpora were also stored for further analysis.

4.4 Data analysis

Corpus linguistics, by definition, can be seen as a discipline in linguistics or a method that investigates a range of linguistic phenomena (McEnery and Hardie, 2012). As a discipline, corpus linguists are often concerned with issues such as the balance and representativeness of a corpus (e.g. Hunston, 2002) or the development and exploitation of a corpus tool (e.g. Hardie, 2012; Anthony, 2023), and these issues are considered "high-level issues of practice" in corpus linguistics (McEnery and Hardie, 2012, pp. xiii). As a method, corpus linguists often use corpus-based or corpus-driven approach to investigate language patterns and variations. McEnery and Hardie (2012) explain that corpus-based studies use corpus data to investigate an existing theory or a hypothesis, whereas corpus-driven studies regard the corpus as "the solely source of

our hypothesis about language" (p.6). They explain that the primary distinction between the two approaches lies in "the degree to which empirical data from a corpus is relied on" (p.151). However, McEnery and Hardie (2012) hold the view that all corpus linguistic studies can be labelled as 'corpus-based', because they argue that the corpus itself does not "[have] theoretical status" (p.6). Based on this argument, the current study is corpus-based, using corpus to investigate how L1 users use PVs and how learners' use of PVs differs from L1 users.

The current corpus-based study integrates both quantitative and qualitative analysis when investigating the use of PVs in the reference corpus and the learner corpus. In corpus-based studies, concordance usually plays a central role in the investigation of language patterns and variations. The term 'concordance', which is also known as KWIC (key words in context), is defined as "all the occurrences of a particular search term in a corpus, presented within the context that they occur in; usually a few words to the left and the right of the search term" (Baker, 2006, p.71). The definition indicates that concordance is the full list of the query, thus corpus linguists usually use the term 'concordance line' to refer to an individual line of occurrence (Gries, 2009; McEnery and Hardie, 2012; Wulff and Baker, 2021). Wulff and Baker (2021, p.165) state that "concordance lines form the basis for both qualitative and quantitative analysis". However, the distinction between quantitative and qualitative analysis seems to be rather vague when it comes to concordance lines because concordance lines are involved in all stages of an analysis (Wulff and Baker, 2021). In other words, concordance lines can indicate both the frequency data of a particular word (i.e. quantitative) and the context where the word appears (i.e. qualitative). As a result, this section aims to discuss the difference between the two types of analysis in detail and

specify what the quantitative analysis involved (see Section 4.4.1) and what the qualitative analysis involved (see Section 4.4.2) in the current study.

4.4.1 Analysing quantitative data

McEnery and Hardie (2012) suggest that quantitative analysis is demonstrated through frequency data. Wulff and Baker (2021) explain that quantitative analysis in corpus linguistics "[focuses] either exclusively, primarily, or in the initial stages of analysis, on the distributional patterns and global statistical trend underlying a given phenomenon" (p.165). Both of the definitions highlight the important role that frequency data plays in relation to a language phenomenon. In the current study, I focused on two types of frequency data. One is the co-occurrence of words (particularly nouns) around specific PVs in the BoE corpus, and the other is the observed (i.e. raw frequency) and normalised frequency of PVs in the reference corpus and the learner corpus.

To answer the first research question, this study focused on the lexico-grammatical patterns associated with specific PVs, such as *LOOK up, LOOK out, LOOK for, LOOK forward to*, etc. They share a same verbal component, *LOOK*, which is one of the most common and frequent verbs in English. As mentioned in Section 3.1.4, the current study follows the frequency-based approach and focuses particularly on the nominal collocates that occur around PVs in the 3:3 span. While researchers have acknowledged the dominant role of frequency in the identification of collocation in the corpus (Stubbs, 2001; Durrant, 2009; Walker, 2011a), some are concerned that regarding observed frequency as the only criterion may be unreliable. For instance, Gablasova, Brezina, & McEnery (2017) illustrated that if a word combination was frequently used by a small

group of language users, it may not be as significant as other frequent combinations used widely by a large number of speakers. For this reason, additional statistical measures of collocation, such as *t*-score and mutual information (MI), are frequently employed in corpus-based studies of collocations (e.g. Durrant and Schmitt, 2009; Granger and Bestgen, 2014; Gablasova, Brezina, & McEnery, 2017). This study identified frequent and significant collocates of PVs using *t*-score. It is a statistical measure used to determine the statistical significance of the co-occurrence of words compared to what would be predicted by chance. If a collocate has a high *t*-score, it generally means that the observed frequency of this collocate in the context of the node is significantly higher than what would be predicted by chance (Gries, 2010). However, MI measure was not adopted because this measure identifies significant collocates with a relatively low frequency.

I paid attention to the second type of frequency data when answering the second and the third research question. The second research question is concerned with the extent to which the use of PVs in the Chinese learner corpus differs from the reference corpus. Table 4.7 shows the observed frequency of the 100 frequent PVs in both corpora. As Table 4.7 shows, there are 9881 occurrences of the 100 PVs in the reference corpus and 5072 occurrences in the Chinese learner corpus. However, many corpus linguists pointed out that the observed frequency of an item in two corpora cannot be compared directly (Gries, 2010; Brezina, 2018). The queried item might appear more frequently in a larger corpus simply because the corpus contains more words overall. In the current study, as mentioned earlier in Section 4.2, the sizes of the reference corpus and the Chinese learner corpus are different. As a result, I also calculated the normalised frequency (i.e. frequency per million words) of the 100 PVs in each corpus. As Table

4.7 shows, the 100 frequent PVs appear 2312.10 times per million words in the reference corpus and 1759.63 times per million words in the Chinese learner corpus.

Table 4.7 The frequency of the 100 frequent PVs in the reference corpus and the Chinese learner corpus

	Observed frequency	Normalised frequency
The reference corpus	9881	2312.10
The Chinese learner corpus	5072	1759.63

Based on the observed and normalised frequency of PVs in the two corpora, it can be seen that the use of PVs differs between the two corpora. However, it may not be a generalisable result. That is, the different use observed in the two corpora may not be a true difference in the population (Oakes, 1998; Brezina, 2018). As a result, in order to assess the difference between the two corpora is not due to chance (i.e. statistically significant), the frequency data was further validated by the chi-square test (χ^2) and *p*-value. Brezina (2018) explained that the chi-square test statistic and the *p*-value can be used to determine whether the difference between the two corpora is significant enough to reject the null hypothesis. In the current study, the hypothesis was that Chinese EFL learners and L1 users use frequent PVs in different ways in written texts. Therefore, the null hypothesis should be there is no difference between the use of frequent PVs in the reference corpus and the learner corpus. As for the levels of significance, if the p-value was less than the significance level (i.e. $p < 0.01$), the null hypothesis was rejected. If $p > 0.05$, then the result failed to reject the null hypothesis, indicating that no significant difference in the use of PVs was observed in the two corpora.

4.4.2 Analysing qualitative data

Besides quantitative data, this study also incorporated qualitative examination of

concordance lines. Many corpus linguists noted that the results of quantitative analysis need interpretation, and the interpretation can be made via qualitative analysis (Wulff and Baker, 2021). Qualitative analysis, according to Wulff and Baker (2021, p.165), "[focuses] either exclusively, primarily or in the initial stage of analysis, on in-depth scrutiny of individual attestations of a phenomenon". I used qualitative analysis in my study for three main reasons. Firstly, it is difficult to rely solely on statistical results to examine the association between the meanings of PVs and the three aspects of collocation, tense and grammatical patterning. Although statistical analysis shows the overall difference regarding the frequency of PVs in the two corpora, qualitative analysis can provide further insights into how Chinese EFL learners use PVs differently compared to L1 users. Secondly, qualitative analysis can reveal nuances that are not easily captured by quantitative methods. This was particularly useful when comparing the use of PVs and their one-word equivalents. For instance, the grammatical pattern where a PV was used might be different from the pattern where its one-word equivalent appeared. This difference in patterning cannot be observed by interpreting the statistical data alone. Thirdly, quantitative analysis cannot explain why Chinese learners use PVs in a particular way. In the current study, I speculated that the use of PVs in learners' writing may be influenced by the way of teaching PVs. Therefore, as part of the qualitative data analysis, I also examined the instruction on PVs in teaching materials for college students in China.

The purpose of examining the instruction in teaching materials was to gain some insights into the way of teaching PVs in the Chinese EFL context and to explore why Chinese EFL learners use PVs in a particular way. As mentioned earlier in Section 4.2.2, the Chinese learner corpus included texts written by undergraduate students in China

and their proficiency level was intermediate. Given the background of the target learners, I chose a course book called *New Horizon College English – Reading and Writing: Student Book*[22] (Zheng, 2015). It is a series of books designed for college students who speak English as a foreign language (Zheng, 2015). I chose this course book mainly because it was the most widely used course book for the College English course[23] in China and has been adopted by numerous universities across the country (Che, Chen and Jiang, 2024). Another reason was that this course book focused on reading and writing skills, which was relevant to the current study. Figure 4.8 below illustrates the way of teaching PVs in the course book. As Figure 4.8 shows, *LOOK back* was regarded as a fixed phrase and students were provided with both English and Chinese explanations. In other words, the instruction on PVs was meaning-oriented. This meaning-oriented technique may cause potential problems because it ignored the contexts in which *LOOK back* was used. Without contexts, students may use PVs in an unnatural way.

Phrases and expressions

be in the / a minority	form less than half of a group 占少数
as for	used for introducing a subject that is related to what you have just been talking about 关于；至于
a matter of (doing) sth.	a situation that involves sth. or depends on sth. 关乎…的事情
look back	think about a time or event in the past 回忆起；回顾
feel sorry for oneself	feel sad about one's life rather than trying to do things that could make one feel better 自我怜悯
be envious of	be wanting sth. that sb. else has 妒忌；羡慕
catch / get / have a glimpse of	have a quick look at sb. or sth. for a moment only 一瞥；飞快地一看
relieve sb. of sth.	(*fml.*) help sb. by taking sth. from them, esp. a job they do not want to do or sth. heavy that they are carrying 解除某人的负担

Figure 4.8 Instruction on *LOOK back* in *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015)

[22] I used the third edition of *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015) in my study. The third edition of this series was published in 2015 and contained four books.
[23] College English is an English language course for undergraduate students from non-English majors in China (Xu and Fan, 2017). This course is primarily oriented to English for general purposes (Xu and Fan, 2017).

To sum up, once the raw data had been collected, it was analysed using both quantitative and qualitative techniques in order to examine how the use of PVs in the Chinese learner corpus differs from that in the reference corpus. Based on the frequency data of PVs in the two corpora, statistical analysis aimed to determine whether the different use of PVs is significantly different in the two corpora. Qualitative examination complemented the statistical results by providing detailed comments on learners' performance when they use PVs in their writing.

4.5 Summary

To sum up, this chapter has described the methodology of the current study, including the corpora used for comparison, the steps of collecting data, and how the data was analysed using quantitative and qualitative techniques. The study aims to answer three research questions. The first research question is concerned with the patterning of PVs in the BoE corpus and the second and the third focus on the difference between the use of PVs in the reference corpus and the Chinese learner corpus. In order to answer the three research questions, this study integrated both quantitative and qualitative analysis. The two types of analysis made complementary contribution to the current study. Quantitative analysis showed the lexico-grammatical patterns associated with PVs in the BoE corpus and presented statistically significant and generalisable results on the different use of PVs in the reference corpus and the learner corpus. Qualitative analysis provided a more precise and in-depth explanation on how individual PVs are used by experienced users of English and identified difficulties or errors of using PVs in learners' writing.

Chapter 5 Using semantic criteria to identify PVs in the corpus

The focus of this chapter is to introduce the semantic criteria used to identify PVs in the current study. This chapter can be regarded as a follow-up to Section 2.2, where I discussed the weaknesses of traditional syntactic tests and the replacement test of PVs. This chapter also serves as part of the methodology for this study, as it provides more details about how I collected and refined data from the corpus. Additionally, it should be noted that corpus examples from the BoE corpus are used for demonstration in this chapter.

Section 5.1 introduces the two semantic criteria and employs corpus data to show how each criterion is used to identify PVs. When dealing with verb + particle combinations in the corpus, it can be observed that these combinations usually have various usages. Therefore, Section 5.2 is concerned with the variation in the use of verb + particle combinations. As corpus evidence shows, a verb + particle combination is possible to be used as a PV in one context and as a non-PV combination in another. Finally, Section 5.3 focuses on verb + particle combinations consisting of three components. It shows that three-word combinations also have both PV and non-PV use in the corpus.

5.1 Semantic criteria of identifying PVs

As mentioned in Section 2.1, combinations that consist of a verb and a particle are highly frequent in English, but not all verbs and particles are combined for the purpose of forming a PV. There are other types of multi-word verbs, such as prepositional verbs, that have fuzzy boundaries with PVs. Syntactic tests, such as the particle movement test and the passivisation test, are not sufficiently reliable to identify all English PVs.

These syntactic tests are mainly designed for transitive PVs and ignore intransitive PVs and inseparable PVs (see Section 2.2.1). In addition to syntactic tests, Bolinger (1971) proposed the replacement test, which focused on the semantic aspect of PVs. Unfortunately, this test is not very successful because many PVs do not have a one-word equivalent (see Section 2.2.2). Although the replacement test fails, it reveals that the meaning of a PV may not be the sum of the meanings of the components. This observation is often considered as one criterion for identifying PVs or idioms (Sinclair, 1996). Therefore, this section begins with the non-compositional meaning of PVs and describes the semantic criteria used in the current study to identify PVs in the corpus.

5.1.1 Meaning non-compositionality

Sinclair (1996, p.79) stated that "sometimes the criterion given for identifying phrasal verbs […] is that the meaning is not the same as the sum of the meaning of the constituent words". Following Sinclair's idea, the first semantic criterion is associated with the non-compositional meaning of PVs. By assessing how the meaning of a verb + particle combination is constructed (i.e. compositional or non-compositional), it should be able to differentiate PVs and non-PV combinations. The following examples from the BoE corpus show the difference between compositional and non-compositional meaning. In the 10 examples, *COME in* has a compositional meaning, whereas *COME across* has a non-compositional meaning.

1	Eventually a policeman	**came in**	to get the woman.
2	The men	**came in**	for dinner, exhausted and ravenous.
3	tomorrow morning you can	**come in**	here and you can explain yourselves to me
4	to see you as soon as you	**came in.**	He said it was urgent.
5	door opened and the counterman	**came in**	with two fresh coffees
6	you remember anything or	**come across**	anything that could be relevant, he said
7	He also	**came across**	a professional actor whose uncanny
8	Well I thought you'd have	**come across**	it before.
9	looking for Michael. I've	**come across**	a book I think would interest him
10	psychological problems they might	**come across**	in their work.

The examination of *COME in* in lines 1 to 5 shows that the meaning of this combination is equal to the meaning of *COME* plus the meaning of *in*. For example, in line 5, it is suggested that 'the counterman' went into the room from the outside. Both the verb and the particle contribute to this interpretation. As a result, it can be argued that *COME in* is a free combination of the verb and the particle because its meaning is composed of *COME* and *in*.

As mentioned in Section 3.1.2, the phraseological approach to multi-word combinations suggests that the components are not restricted by each other in free combinations (Cowie, 1988; Howarth, 1996). The substitution of one component does not affect the meaning of the other. Consider the following substitutions of line 5:

The counterman ***came out*** with two fresh coffees.

The counterman ***went in*** with two fresh coffees.

In the first substitution, the meaning of *came* does not change when *in* is substituted by *out*. The meaning of the verbal component remains to indicate 'the motion in space' (*OED online*). Likewise, replacing *came* with *went* does not affect the meaning of *in*. According to the phraseological approach proposed by Cowie (1988), it can be argued that the combination of *COME* and *in* in 'the counterman *came in* with two fresh coffees' can be seen as a free combination. Another observation about the use of *COME in* is that

the verbal component preserves its meaning even without the non-verbal component.

The counterman ***came*** (*in*) with two fresh coffees.

As the example shows, there is no significant change in the meaning of this sentence if *in* is deleted. It seems that the function of *in* is to add a layer of meaning to *came*, explaining the direction of the motion (i.e. from outside to inside).

In terms of *COME across*, the examination of lines 6 to 10 suggests that the way of interpreting *COME in* does not work to interpret *COME across* properly. In other words, the meaning of *COME across* in the given examples cannot be predicted by the two components. The meaning of *COME across* in lines 6 to 10 seems to be associated with 'meeting someone or encountering something'. For example, in line 7, the contextual information suggests that the man met a professional actor. In line 10, it is suggested that they might encounter psychological problems in their work. The interpretation of *COME across* has no association with either the verbal component or the non-verbal component. This argument can be supported by the following instances.

*He also ***came*** a professional actor.

*Psychological problems they might ***come*** in their work.

As the two instances show, if the particle is ignored, the individual verb *came* makes no sense in the context. Also, if other verbs substitute the verbal component in the combination, the meaning of *across* will be influenced.

```
11              We had to get across the river before they arrived at the bank
12               two kids ran across the road and into the dark
13          One sailed slowly across the room, hit the carpet
14         clouds slid slowly across the sky
15                 She walked across the park towards City Road and King Street
```

In lines 11 to 15, the verbal component has changed. These verbs, such as *GET, RUN,*

and *WALK*, share a semantic aspect with *COME*, suggesting the motion from one position to another. When *across* follows these verbs, the meaning of this particle can be interpreted as 'from one side to the other' (*OED* online). For example, in line 12, 'two kids ran across the road' suggests the quick movement from one side of the road to the other side.

Based on the above analysis of *COME across*, it can be argued that both components are mandatory in the combination, although the meanings of the two components do not contribute to the understanding of the combination. The substitution or change of one component impairs the meaning of the other. In addition, the comparison between *COME across* and other combinations such as *RUN across* and *WALK across* indicates that the combination of *COME* and *across* is not free and the meaning of this combination is non-compositional. The current study classified *COME across* as a PV, whereas other combinations such as *GET across, RUN across, WALK across*, etc. in the given concordance lines were non-PV combinations of a verb and a particle. The main criterion that is used to distinguish PVs from non-PV combinations is whether the meaning of the combination is compositional. If the verb and the particle are combined in a way as free combinations and both play a part in predicting the meaning of the combination, the combination will be classified as a non-PV combination in the current study. If, on the other hand, a verb + particle combination has a non-compositional meaning (i.e. its meaning cannot be predicted by the meanings of the components), it will be regarded as a PV in the current study.

It should be noted that the classification of PVs and non-PV combinations is closely related to the context in which verb + particle combinations appear. In the given

concordance lines (i.e. lines 11 and 12), combinations such as *GET across* and *RUN across* are non-PV combinations because their meanings are compositional and predictable. However, this does not necessarily mean that these combinations cannot be PVs. In the two corpus examples below, the meanings of *GET across* and *RUN across* seem to be non-compositional. I will return to this point in Section 5.2 which focuses on various usages of verb + particle combinations.

> This is a powerful and flexible way to ***get*** information ***across***. (BoE: BB_Qm90_317)
>
> You'd be lucky to ***run across*** him in a year. (BoE: BA_aM032570)

In the first stage of data collection (see Section 4.3.1), the corpus tools retrieved concordance lines that contained verb + particle combinations, but the computer could not distinguish which combinations were PVs and which were not. Using this semantic criterion, it was possible to manually identify and filter PVs from the large collection of verb + particle combinations in the corpus. The following concordance lines exhibit more examples of PVs. In these lines, each verb + particle combination has a non-compositional meaning.

16		I had to	**give up** my job as a secretary.
17		but he refused to	**give up** and was back on the road again
18		He would never	**give in** and never take things easy.
19		I don't think he ever	**got over** the death of his parents, losing them
20	particularly the blacks who		**make up** 10 per cent of the US population.
21		I may	**take after** my father in every respect except
22	spending hours at the gym		**working out**, or going to an aerobics class
23	it was he and his kind who		**brought** these conditions **about** in South Africa,
24		I've just	**taken** it **up** again after a break
25	his foreplay, which		**turned** him **off** and caused him to lose interest.

Compared with the syntactic tests that fail to account for inseparable PVs and intransitive PVs (see Section 2.2.1), the semantic criterion that examines meaning

compositionality seems to be more inclusive. Intransitive PVs, such as *GIVE up* (line 17), *GIVE in* (line 18) and *WORK out* (line 22), can be identified because all of them have non-compositional meanings in the given concordance lines. Inseparable PVs, such as *COME across* (lines 6 to 10), *GET over* (line 19) and *TAKE after* (line 21), can be identifiable with the help of this semantic criterion. Moreover, it also identifies PVs that are separated by intervening items (e.g. lines 23 to 25).

5.1.2 Meaning extension of the component

Section 5.1.1 has introduced one semantic criterion used in the current study. If a verb + particle combination has a non-compositional meaning, it will be regarded as a PV in the current study. In addition to PVs, the first semantic criterion can be used to identify idioms because idioms also have unpredictable meanings (Sinclair, 1996; Moon, 1998). However, in contrast to idioms, Sinclair (1996) noted that this criterion cannot be a formal criterion for PVs because the individual words within a PV may have other sorts of relationships to the overall meanings. In other words, non-compositional meaning is only one of the semantic features of PVs. Among the large collection of verb + particle combinations in the corpus, it is possible to find some PVs whose meanings are partially predictable. The meaning of the combination seems to have some associations with one of the components. For example, the following corpus evidence shows that the meaning of *TURN on* is partially predictable and closely associated with the non-verbal component *on*.

26		lock the door and	**turn on** my hair dryer
27	went into the bedroom, while I		**turned on** the TV and watched mindless pop videos
28	He started the motor and		**turned on** the radio to find the station
29		I	**turned** a light **on**, stumbled to the door
30	waiter came over and		**turned** the computer **on** for him.

In lines 26 to 30, the objects that can be turned on are electronic devices, such as 'hair

dryer', 'TV', 'radio', 'light', and 'computer'. The meaning of *TURN on* in the five corpus examples can be interpreted as 'to power up the electronic devices by pressing the switch or the button'. It is plausible to claim that *TURN on* has a non-compositional meaning because the meaning of the verb and the spatial meaning of the particle are not associated with the interpretation. However, if the meaning of *on* in the following examples is considered, the meaning of *TURN on* seems to be compositional and predictable.

31	he was aware of the microphone being	**on**	or not, Ron Atkinson has shown
32	When your computer is	**on**	but not in use
33	but the radio was	**on**	and that beautiful voice was
34	The TV was	**on**	but silent.
35	The lights were	**on**	and the theatre was empty.

Lines 31 to 35 provide corpus evidence to demonstrate the extended meaning of the particle *on*. As the five lines show, *on* is sometimes used as an adjective, describing 'the state of an electronic device' (*OED online*). For example, in line 32, the use of *on* indicates that the computer is powered up and ready for use. This observation suggests that the meaning of *on* in *TURN on* is similar to the meaning in the given examples (i.e. lines 31 to 35). Since the extended meaning of the particle (i.e. the adjective meaning of *on*) helps to predict the meaning of *TURN on*, we cannot classify this combination as a PV using the first semantic criterion. However, there is no corpus evidence showing the meaning connection between *TURN on* and the individual verb *TURN*. For instance,

36	He hung up and	**turned**	around to find Lisa
37	Now I see the wheel	**turning**	full circle.
38	Arnold	**turned**	his head and looked at the kid
39	street of the city, he	**turned**	left and disappeared into
40	I	**turned**	the handle and found the door was open.

When *TURN* is not followed by *on*, lines 36 to 40 illustrate the frequent use of this verb in the BoE corpus. The examination of these concordance lines indicates that the verb usually describes an action of moving the body (e.g. lines 36 and 38), rotating (lines 37

and 40), or moving towards a direction (line 39). Compared with the meaning of *TURN* in the five examples, it seems that the verb has lost its literal meaning when combined with *on*. The meaning of *TURN* becomes delexical[24] when it is combined with *on* in the example.

Based on the analysis of *TURN on*, it is possible to introduce the second semantic criterion: if one component in a verb + particle combination is used in its extended meaning, the verb + particle combination can be regarded as a PV. It should be noted that this semantic criterion often identifies verb + particle combinations whose meanings are partially compositional. More specifically, the extended meaning of *on* contributes to understanding the meaning of *TURN on*, whereas the meaning of *TURN* does not. When one component in the PV has an extended meaning, the meaning of the other component tends to be weak and less central. As the corpus evidence of *TURN on* shows, the verbal component loses its inherent meaning and has a weak association with the action of moving or rotating.

In addition to *TURN on*, another verb + particle combination, *LOOK into*, can be regarded as a PV using the second semantic criterion. The combination of *LOOK* and *into* could be identified as a PV in the following concordance lines.

41		We're	**looking into** the death of Mr Philip Kershaw
42		You might also	**look into** the possibility of buying a good
43	He may be lying.	We're	**looking into** the situation. If his information
44	have a responsibility to		**look into** those complaints on behalf of the fans
45		they'll go and	**look into** whatever is going on

In the five corpus examples, the meaning of *LOOK into* can be interpreted as 'to examine

[24] 'Delexical' usually refers to a verb that has little or no inherent meaning by itself but gains meaning in combination with other words, such as nouns or adjectives (Hunston, 2002). When a verb is delexicalised in a word combination, its meaning becomes weak and less central.

or investigate in order to gather more information'. For example, in the context of line 43, 'the situation' is being investigated in order to confirm whether 'he' is lying or not. When interpreting the meaning of *LOOK into* in the given corpus data, it seems that the figurative meaning of *into* in lines 46 to 50 contributes to this interpretation.

46	powers to investigate complaints	**into**	the activities of garda in Donegal
47	to raise funds for research	**into**	the causes and prevention of SIDS
48	no formal inquiry	**into**	the cause of the disaster was done
49	in charge of a major investigation	**into**	the murder of several GRU officers
50	published the results of a study	**into**	the psychiatric state of

As a result, using the second semantic criterion, *LOOK into* can be classified as a PV in lines 41 to 45 because the meaning of the particle has been extended. Meanwhile, the verb *LOOK* loses its literal meaning, having no association with 'to use one's eyes to see' (*OED online*). Researchers who identify PVs by the particle movement test proposed by Bolinger (1971) may ignore the combination of *LOOK* and *into* and regard it as a non-PV combination. Using the traditional syntactic tests (see Section 2.2.1), *LOOK into* may not be classified as a PV because it does not allow the separation of the components (**looking* the death *into* or **look* the possibility *into*). However, the argument that *LOOK into* should be classified as a PV disagrees with the traditional view that regards this combination as a prepositional verb. In the current study, *LOOK into* was classified as a PV when it was used in examples such as '*look into* the possibility' (line 42) and '*looking into* the situation' (line 43).

Using the second semantic criterion, it would be reasonable to reject *STAND up* and *SIT down* as PVs in the following concordance lines.

51 ask our new companions to **stand up** and introduce themselves
52 He **stood up** and left the reading room
53 He **stood up** and walked to the door of his office
54 The officer **stood up** as Malone and Gail came in.
55 finished his beer quickly and **stood up** to leave.

56 found a chair and **sat down**. He was silent for a long
57 to the group and they **sat down** and talked for a while.
58 I **sat down** in front of my computer screen
59 a place where we can **sit down** and have something to drink.
60 my mother said, **sitting down** beside me on my bed,

The examination of *STAND up* in lines 51 to 55 suggests that both the verbal component and the non-verbal component are used in their literal meanings, without any meaning extension observed in the five examples. This is also true with respect to *SIT down* in lines 56 to 60. Using the second semantic criterion, the current study did not classify *STAND up* and *SIT down* as PVs in the given contexts. However, lines 61 to 65 provide corpus evidence to show that experienced users of the language sometimes use the combination of *STAND* and *up* in a different way.

61 How well does this thesis **stand up** to close examination?
62 the chances are they would not **stand up** to close scrutiny.
63 personal insults and less able to **stand up** to injustices.
64 because they cannot **stand up** to the harsh conditions
65 No bulkhead is going to **stand up** to the pressure at full speed

More precisely, the five examples illustrate the use of *STAND up to*, a three-word combination. It can be observed that *STAND up* is consistently followed by *to* in the five examples, and the sentences do not make sense if *to* is absent.

*How well does this thesis ***stand up*** close examination?

In the current study, the combination *STAND up to* in lines 61 to 65 was classified as a PV because the verbal component was used in its extended meaning. For example, the meaning of *STAND up to* in line 61 and 62 seems to be associated with the ability of resisting or withstanding examinations or scrutiny. This extended meaning of *STAND* can be demonstrated by the following examples.

66 and I can't **stand** any criticism of the way I'm doing it
67 We can **stand** it for a lot more than a year
68 how long the body can **stand** pressure like that,
69 but Marianne could not **stand** the isolation and ran away.
70 felt his heart could not **stand** the strain of a flight

The meaning of *STAND* in lines 66 to 70 is associated with 'to confront or endure' (*OED online*), which could help to understand the meaning of *STAND up to*. Unlike *TURN on* and *LOOK into* in which the particles have extended meanings, the combination *STAND up to* is an example in which the verbal component has an extended meaning. In addition to two-word PVs, three-word PVs, such as *STAND up to*, can also be identified using the second semantic criterion. Section 5.3 will provide more details about the identification of three-word PVs using the semantic criteria.

Compared with traditional syntactic tests (e.g. the particle movement test), using the second semantic criterion to examine meaning extension of the component can identify both separable PVs (e.g. *TURN on*) and inseparable PVs (e.g. *LOOK into*). This criterion is often applicable to verb + particle combinations whose meanings are partially compositional. That is, the extended meaning of one component helps to predict the meaning of the combination, while the other component loses its literal meaning in the combination. However, verb + particle combinations are frequently used in multiple ways in the corpus (e.g. 'he *stood up* and left the reading room' vs 'they cannot *stand up to* the harsh conditions'). This suggests that using semantic criteria to identify PVs depends on the actual use of verb + particle combinations in the context. Therefore, the next section (Section 5.2) will focus on the variation in the use of verb + particle combinations.

5.2 The variation in the use of verb + particle combinations

In Section 5.1, I have introduced two semantic criteria that were used to identify and classify PVs in the current study. It should be noted that the two semantic criteria are not intended to be exhaustive due to the various ways in which verb + particle combinations can be used. A verb + particle combination can be used as a PV in one context and as a non-PV combination in another. For example,

71 The train **came across** the bridge over the dry lake
72 as Gertrustein and Sam **came across** the garden, and Anastasia ran to open
73 holding his breath as he **came across** the room. Cat moved to
74 were closed against the wind **coming across** the bay, turning the water into
75 He was **coming across** the river with some woman

In contrast to the use of *COME across* in 'he *came across* a professional actor' (line 7), the combination of *COME* and *across* was classified as a non-PV combination in lines 71 to 75. In line 71, for example, the meaning of '*came across* the bridge' is associated with moving from one side of the bridge to the other, which can be predicted by the meanings of the two components. In other words, *COME across* has a compositional meaning, indicating that this combination cannot be classified as a PV in lines 71 to 75. As a result, the combination *COME across* has various usages in the corpus. It is classified as a PV in 'he *came across* a professional actor' because of the non-compositional meaning indicated by the context, whereas it is classified as a non-PV combination in 'the train *came across* the bridge' where it has a compositional meaning. This is also true with another verb + particle combination, *RUN across*. The use of *RUN across* in line 12 (i.e. as a non-PV combination) is different from the use in lines 76 to 80 (i.e. as a PV).

76 the police **ran across** a fisherman in Miminegash
77 I **run across** a lot of people who say
78 you'd be lucky to **run across** him in a year
79 mature young pitchers I've **run across** in a long time
80 You **run across** some object that you haven't seen

When *COME across* and *RUN across* are classified as non-PV combinations, it seems that the corpus evidence demonstrates the grammatical patterning of these verbs. The grammatical pattern is that the verb is followed by *across* and a noun phrase (i.e. **V *across* n**).

11	We had to **get**	**across**	the river before they arrived at the bank
12	two kids **ran**	**across**	the road and into the dark
13	One **sailed** slowly	**across**	the room, hit the carpet
14	clouds **slid** slowly	**across**	the sky
15	She **walked**	**across**	the park towards City Road and King Street
71	The train **came**	**across**	the bridge over the dry lake
72	as Gertrustein and Sam **came**	**across**	the garden, and Anastasia ran to open
73	holding his breath as he **came**	**across**	the room. Cat moved to
74	closed against the wind **coming**	**across**	the bay, turning the water into
75	He was **coming**	**across**	the river with some woman

As the corpus evidence shows, verbs with this pattern consist of *COME, RUN, WALK,* etc. The shared meaning aspect of these verbs is concerned with "crossing or passing from one side of a place to the other" (Francis, Hunston, and Manning, 1996, p.151).

When *COME across* and *RUN across* are classified as PVs, their meanings tend to be associated with 'meeting someone' (e.g. lines 7, 76, 77, 78, and 79) or 'finding something' (lines 6, 8, 9, 10, and 80). Corpus evidence suggests that *COME across* co-occurs more frequently with noun phrases such as 'a book', 'psychological problems', etc. However, noun phrases that frequently follow *RUN across* tend to refer to a person, such as 'a fisherman' and 'young pitchers' .

6	you remember anything or	**come across**	anything that could be relevant, he said
7	He also	**came across**	a professional actor whose uncanny
8	Well I thought you'd have	**come across**	it before.
9	looking for Michael. I've	**come across**	a book I think would interest him
10	psychological problems they might	**come across**	in their work.
76	the police	**ran across**	a fisherman in Miminegash
77	I	**run across**	a lot of people who say
78	you'd be lucky to	**run across**	him in a year
79	mature young pitchers I've	**run across**	in a long time
80	You	**run across**	some object that you haven't seen

Moreover, replacing the verb with one that shares a similar meaning is not permissible in PVs. For example, the verb *GET* shares a meaning aspect with *COME* and *RUN* in '*get across* the river' (line 11), but it does not form a PV in a same way as *COME* and *RUN*.

81 What they're trying to **get across** to you is that for security reasons
81 One way of **getting** information **across** to people would be
83 our point of view and **getting** it **across** to the organisers.
84 in what you're doing and **getting** that **across** to the audience.
85 the most effective way of **getting** the message **across**.

Using the first semantic criterion, the combination of *GET* and *across* in lines 81 to 85 was classified as a PV and the meaning of *GET across* was associated with 'conveying a message or a piece of information so as to be understood'. However, it is impossible to replace *GET* with *COME* and *RUN* in these examples.

One way of ***getting*** information ***across*** to people would be

*One way of ***coming*** information ***across*** to people would be

*One way of ***running*** information ***across*** to people would be

The above analysis has shown that PVs which are identified using the first semantic criterion (e.g. *COME across, RUN across,* and *GET across*) are possible to be non-PV combinations. In terms of the PVs whose meanings are partially compositional, they are also possible to be non-PVs combinations. For instance,

86 I **looked into** her dark eyes and saw
87 Gana stooped and **looked into** the baby's face, touching it
88 Now she **looked into** the mirror and smiled.
89 As I **looked into** the nest, I saw that the mother bird
90 open doorways, she **look into** the room. The moonlight shining

When understanding the meaning of *LOOK into* in lines 86 to 90, it is unnecessary to examine the extended meaning of *into*, as in 'the research *into* the causes' (line 47). The two components contribute to predicting the meaning of *LOOK into*, but neither has an extended meaning. As a result, the combination of *LOOK* and *into* is classified as a non-PV combination in lines 86 to 90 because it has a compositional meaning.

In this section, I have examined the variation in the use of verb + particle combinations. Corpus evidence reveals that these combinations are often used in various ways, so the identification of PVs has to be determined by the context. Without contextual information, it should be cautious to determine whether a verb + particle combination is or is not a PV. Based on the examination, it seems that the variation in the use of verb + particle combinations can be roughly divided into non-PV usage (e.g. '*look into* the mirror) and PV usage (e.g. '*look into* the possibility'). In the non-PV usage, the verb + particle combination has a compositional meaning, with the literal meaning of each component being preserved. On the other hand, in the PV usage, the meaning of the verb + particle combination tends to be non-compositional (i.e. unpredictable by the components) or partially compositional (i.e. associated with the extended meaning of one component while the other component loses its literal meaning in the combination). This observation indicates that the notion of PVs is closely connected with what Cowie calls figurative idioms. As mentioned in Section 3.1.2, figurative idioms refer to combinations which has figurative meanings but also allow for literal interpretations (Cowie, 1981; Howarth, 1996).

When verb + particle combinations are used as PVs, the examples illustrated in this section, such as *COME across, RUN across, LOOK into,* etc., do not exhibit multiple meanings. More specifically, as PVs, their meanings are less varied. For instance, when *LOOK into* is used as a PV, its meaning is associated with 'to examine or investigate in order to gather more information'. However, it should be noted that this may not be true for all PVs in English. One of the typical features of English PVs is that they have multiple meanings (see Section 2.1.3). It is the polysemous nature of PVs that poses

great challenges for ESL/EFL learners (Sinclair, 1996; Siyanova and Schmitt, 2007). Chapter 6 and 7 will look into more verb + particle combinations, such as *LOOK up*, *LOOK for*, *LOOK forward to*, etc., whose meanings tend to be more complex when they are used as PVs.

5.3 The identification of three-word PVs

In the last two sections, I have demonstrated the process of using the semantic criteria to identify and classify several two-word PVs, such as *COME across, TURN on, LOOK into*, etc. The final issue this chapter needs to address is the identification of three-word PVs. Conventionally, three-word combinations, such as *CATCH up with, LOOK forward to, PUT up with*, etc. are called phrasal-prepositional verbs (Quirk et al., 1985; Biber et al., 1999; Downing and Locke, 2002). This traditional label focuses on the grammatical function of each component but tends to be quite arbitrary because there is no specific criterion explaining why the second component is adverbial and the third is prepositional. Celce-Murcia and Larsen-Freeman (1999) regarded these three-word combinations as three-word PVs because they are similar to two-word PVs, functioning as single lexical units. Based on Celce-Murcia and Larsen-Freeman's idea, this section aims to examine whether the semantic criteria are also applicable to three-word PVs and investigate the variation in the use of these three-word combinations using corpus data. Here are some corpus data showing the use of a three-word combination, *PUT up with*.

```
91              team managers have had to put up with extraordinary pressure and stress
92                  those who have had to put up with her pain and suffering over
93                   I love him but can't put up with his computer addiction
94    weather conditions that you have to put up with or adapt to.
95         become angry at themselves for putting up with unfair demands.
```

Based on the contextual information around *PUT up with* in lines 91 to 95, this

combination co-occurs frequently with expressions implying a negative meaning, such as ‘extraordinary pressure and stress’ (line 91), ‘her pain’ (line 92), ‘unfair demands’ (line 95). However, the three components of this combination do not have negative meanings or imply negative connotations. This suggests that the meaning of *PUT up with* cannot be predicted by the components. Using the first semantic criterion, *PUT up with* is classified as a PV in the given concordance lines because its meaning is non-compositional.

In addition to *PUT up with*, the following corpus data for *LOOK down on* indicates that the second semantic criterion is also applicable.

96		some of them used to	**look down on** my dad because he didn't work
97	Chinese traditional society always		**look down on** the people who have money.
98		Nobody's going to	**look down on** you because you work here
99		urban slums, they were	**looked down on** by most native-born Americans
100	aristocrats who have previously		**looked down on** her as upper middle class

It appears that *LOOK down on* has a partially compositional meaning in the context because the extended meaning of *down* helps to understand this combination. The extended meaning seems to be associated with the metaphor “HIGH STATUS is UP; LOW STATUS is DOWN” (Lakoff and Johnson, 1980, p.16). In the given examples, the different social status is implied by *LOOK down on*. For example, line 100 implies the difference between ‘aristocrats’ and ‘upper middle class’ using *LOOK down on*. The ‘upper middle class’ is perceived to be in a lower position than ‘aristocrats’. *LOOK down on* is therefore classified as a PV in lines 96 to 100 using the second semantic criterion. However, if we consider the following concordance lines where *LOOK down on* also appears, it is not possible to classify this combination as a PV in lines 101 to 105.

101 From the top road, you **look down on** the houses and gardens below
102 so you can **look down on** the river from above
103 next time you get to **look down on** traffic from a tall building
104 When I **looked down on** it from my new windows
105 A woman **looked down on** the scene from her balcony.

In the five examples, the orientation from upwards to downwards can be indicated by *LOOK down on*. The meaning of this combination becomes compositional, and each component plays a part in understanding this up-down orientation.

Similar to *LOOK down on*, *GET out of* also has both PV and non-PV uses in the corpus. Lines 106 to 115 below can be divided into two groups. The first group (lines 106 to 110) demonstrates the non-PV use of this combination, whereas the second group (lines 111 to 115) shows the PV use of this combination.

106 they like to **get out of** the city and go windsurfing
107 She just wanted to **get out of** there, get home
108 For example, **getting out of** bed, having breakfast,
109 Ken **got out of** the building they were keeping him in
110 our employee **got out of** the vehicle to have a look

111 and you don't **get out of** a long-term problem in the short term
112 so I can **get out of** boring conversations easily
113 asking for help to **get out of** the situation he was in
114 were able to **get out of** their obligation to serve
115 his government is **getting out of** the business of subsidizing

In lines 106 to 110, *GET out of* is used as a non-PV combination whose meaning is associated with leaving or going away from a place such as 'the city', 'the building', 'the vehicle', etc. It has a compositional meaning in the context and no component is used in an extended way. Also, the examination of this combination shows that substituting one component does not influence the meaning of the rest. For instance,

Ken ***got out of*** the building (line 109)

→ Ken ***went out of*** the building.

However, in lines 111 to 115, there is no physical place indicated in the context. For

instance, '*get out of* boring conversations' in line 112 suggests that the person intends to avoid boring conversations. In line 115, a possible interpretation of '*getting out of* the business of subsidizing' could be that the government intends to stop such business. Based on the interpretations, the meaning of *GET out of* should be understood in a figurative way, indicating an action to avoid or stop doing something. In this figurative meaning, it seems that *out* is used in an extended way, contributing to understanding the figurative meaning in lines 111 to 115.

As a result, the examples of *PUT up with*, *LOOK down on* and *GET out of* have shown that the two semantic criteria also work to identify three-word PVs. Similar to two-word combinations, three-word combinations also have both PV (e.g. '*looked down on* my dad because he didn't work') and non-PV use (e.g. '*look down on* the river from above'). Using semantic criteria to classify the use of three-word combinations also depends on the context in which the combination appears.

5.4 Summary

To sum up, this chapter has introduced the two semantic criteria and employed a large amount of corpus data to demonstrate how these two criteria are used to differentiate PVs and non-PV combinations in the current study. When dealing with verb + particle combinations, the semantic criteria focus on how the meaning of the combination is established (i.e. compositional, partially compositional, or non-compositional), rather than anticipating the grammatical function of each component (i.e. adverbial or prepositional) in the combination.

Compared with traditional syntactic tests that often overlook inseparable PVs and intransitive PVs, the semantic criteria appear to be more inclusive because they can identify both types of PVs. For example, intransitive PVs, such as *WORK out*, can be identified in the context 'spending hours at the gym *working out*' (line 20) using the first semantic criterion. Also, inseparable PVs, such as *COME across*, can be identified in the context '*came across* a professional actor' (line 7) using the second semantic criterion. In addition, the semantic criteria are also applicable to three-word PVs (e.g. *PUT up with, LOOK down on,* etc.).

It should be noted that verb + particle combinations usually have various usages (see Section 5.2). This observation suggests that PVs such as *COME across, LOOK into*, etc. could be considered figurative idioms in the current study because they can be used either as PVs or as non-PV combinations. However, unlike the phraseological approach, which relies on researchers' intuition to distinguish between the literal and figurative uses of figurative idioms, the current study emphasizes the critical role that the context plays. Using semantic criteria to examine whether a combination is used as a PV or a non-PV combination depends significantly on the context. It is not possible to classify *LOOK into* as a PV without specifying the context in which it occurs. While the label 'phrasal verbs' is used in this thesis for simplicity reasons, it should be noted that the literal use of PVs (i.e. when they are used as non-PV combinations) is also considered in the current study.

Chapter 6 How do collocation and tense forms affect the meaning of PVs

In the previous chapter, I have discussed how to use semantic criteria to differentiate PVs and non-PV combinations. Based on the discussion in Section 5.1, English PVs generally have two semantic features: (1) the meaning of a PV is non-compositional; (2) the meaning of one component in a PV should be understood in a metaphorical or figurative way and the meaning of the other component gets weak in the combination.

Also, as Section 5.2 shows, the various usages of a verb + particle combination is largely dependent on the context where it appears (e.g. '*COME across* the road' vs. '*COME across* a professional actor'). More specifically, it is not the combination itself that determines the meaning. The lexico-grammatical patterns associated with verb + particle combinations, such as collocation, grammatical patterns and choices of tense forms, may impact their meanings and influence whether a combination should be classified as a PV or a non-PV combination. In this chapter, Section 6.1 will focus on how collocation impacts the meaning of verb + particle combinations. Section 6.2 is concerned with the choices of the progressive form and investigates the influence of this specific form on the meanings of PVs.

6.1 Using collocation to identify various meanings and usages of PVs

The combination of *LOOK* and *up* was selected to demonstrate how collocation works to influence the meaning. This combination occured 17709 times in the BoE corpus. Considering the challenge of analysing the larger number of concordance lines qualitatively, 500 random concordance lines in the BoE corpus were selected for a close examination. The examination of these 500 random samples showed that the contiguous

form of this separable PV (i.e. *LOOK* and *up* are combined without interruption) was much more frequent than the non-contiguous form. The majority of the concordance lines (463 lines) demonstrated the contiguous form of the combination, whereas only 37 lines showed the non-contiguous form. However, it remained problematic whether the contiguous use of *LOOK up* and the non-contiguous use of *LOOK* + *up* conveyed a consistent meaning in the corpus. This section examines *LOOK up* and *LOOK* + *up* individually and presents the findings on the nominal collocates associated with each type.

6.1.1 *Look up*: collocation and meanings

When *up* followed *LOOK* contiguously, Table 6.1 lists ten most frequent nominal collocates that appeared on the left and right side of *LOOK up* in the BoE corpus. The word span on both sides was three words.

Table 6.1 Ten most frequent nominal collocates of *LOOK up* in the BoE corpus: *t*-score and frequency

Freq.	*t*-score	Collocate	Node	Collocate	*t*-score	Freq.
401	18.45	thing		sky	15.74	253
105	9.14	eye		eye	10.74	137
167	7.95	people		face	9.63	128
96	7.76	head		ceiling	9.29	88
122	7.21	man	***LOOK up***	word	8.78	100
63	6.84	moment		star	6.81	67
70	6.09	face		surprise	6.80	53
66	5.84	hand		scoreboard	6.22	39
48	5.81	guy		book	6.18	64
46	5.70	kid		number	6.09	70

As Table 6.1 shows, *eye* was one of most frequent collocates that appeared on both sides of *LOOK up*. The high frequency of *eye* around the target combination suggested that the

meaning of *LOOK up* had a strong association with the literal sense of *LOOK* (i.e. 'to use eyes to see'). The following 10 concordance lines provid evidence for this association.

1	Klessmann opened his **eyes** and	**looked up**	at her, the hint of a smile
2	Katren's soft blue **eyes**	**looked up**	at Brockwell full of sympathy
3	his **eyes** when he	**looked up**	and smiled at me
4	his shaky hands, he	**looked up**	into her **eyes** as if confused by
5	lifted her head and	**looked up**	into my **eyes** slowly and without emotion
6	She	**looked up**	into his **eyes** and saw that
7	Gretchen read her letter and	**looked up**	with excited **eyes.**
8	And when she	**looks up**	with shining **eyes** and
9	At that second, the girl	**looked up**	with frightened **eyes** and
10	the others	**looked up**	with **eyes** dark and dreadful

The patterns associated with *LOOK up* can be observed in the ten concordance lines, especially when *eye* appeared at the third position on the right side of the node (i.e. N+3). When *eye* appeared at N+3, the frequent patterns tended to be '*LOOK up* + into + possessive adjective + eyes' (lines 4 to 6) and '*LOOK up* + with + adjective + eyes' (lines 7 to 10). In the BoE corpus, there were 20 concordance lines showing the first pattern and 24 showing the second pattern. The two patterns also indicated the frequent grammatical choices and semantic preferences associated with *LOOK up*. For example, when the preposition *into* appeared at N+1, L1 users tended to employ possessive adjectives at N+3 (e.g. *my*, *her*, *his*). When the preposition *with* appeared at N+1, the pattern usually included an adjective related to mood at N+2 or nearby (e.g. *excited*, *frightened*, *dark and dreadful*). In other words, when *LOOK up* co-occurred with *eye*, it had a semantic preference for words of mood (see lines 7 to 10). However, it was not evident that the collocation had a negative or positive semantic prosody because both negative and positive expressions were possible to appear.

Besides the collocate *eye*, Table 6.1 also shows that words such as *sky, ceiling* and *star* frequently appeared on the right side of *LOOK up.* The three nominal collocates shared a common feature that they all referred to space or something visible in a very high

position. Therefore, the frequent occurrences of the three collocates were probably associated with the literal meaning of *up* in the combination. Lines 11 to 15 below illustrate the association between *LOOK up* and *sky, ceiling,* and *star*. As the five concordance lines show, when *LOOK up* was followed by collocates such as *sky, ceiling,* and *star*, the frequent pattern can be written as '*LOOK up* at the sky/ceiling/stars'.

11	I was lying on the sidewalk	**looking up**	at the **sky**.
12	John	**looked up**	into the **sky** for a minute.
13	with his arms behind his head,	**looking up**	at the **ceiling**, upon which lights
14	Jacopo lay on his back and	**looked up**	at the **stars**.
15	On romantic nights, when we	**looked up**	at the **stars**, we were looking through telescopes

The examination of the collocation associated with *LOOK up*, including '*LOOK up* into one's eyes', '*LOOK up* with + mood-associated adjective + eyes', and '*LOOK up* at the sky/ceiling/stars', showed a tendency for this combination to have a compositional meaning. The compositional meaning, labelled as *LOOK up*$_1$ in this thesis, can be predicted by the literal meanings of *LOOK* and *up*, which was associated with the action of using or raising eyes physically in order to see something. Moreover, because of the first semantic criterion for PVs mentioned in Section 5.1, the combination of *LOOK* and *up* cannot be classified as a PV when *LOOK up*$_1$ made sense in the context.

In Table 6.1, it is worth noting that *thing* was the most frequent nominal collocate that typically appeared on the left side of *LOOK up.* The high *t*-score of *thing* also indicated the significant association between this collocate and the node. The following 10 examples from the BoE corpus exemplified the co-occurrence of *thing* and *LOOK up*.

16 about his father, **things** are **looking up** for him. He thinks it's better
17 Good news is that **things** are **looking up**. He will be back.
18 have to admit, **things** are really **looking up** for me since I started on the Zoloft.
19 got better and **things** are also **looking up** for this year.
20 disappointed, but **things** started **looking up** after we decided to widen our search
21 manager, believes that **things** are **looking up** for the club, but they could be back
22 a good sign that **things** could be **looking up** in the computer industry.
23 is confident that **things** will **look up** once the company broadens its product portfolio
24 cookery school, **things** begin to **look up**. We are shown to large rooms and
25 **Things** started to **look up** last year when Jade's mum Kate spotted the ad

At first glance, when *thing* co-occurred with *LOOK up*, it was typically used in its plural form, *things*. The examination of lines 16 to 25 showed that *LOOK up*$_1$ no longer made sense when *things* appeared in front of the node. There was no evidence showing the relationship between using eyes and the collocation 'things + *LOOK up*'. However, it was possible to observe a new pattern associated with *LOOK up* when it co-occurred with *things*. When *things* appeared at the N-2 or N-3 position, the node had frequent grammatical choices. For instance, it was used in the progressive tense (e.g. 'are *looking up*' in lines 16 and 17). In some cases, it followed expressions indicating the future (e.g. 'will *look up*' in line 23 and 'begin to *look up*' in line 24). The grammatical choices suggested that the meaning of *LOOK up* tended to be related to something that is changing or will change in the future. Moreover, a closer examination of 'things + *LOOK up*' in lines 16 to 25 showed that this collocation tended to co-occur with expressions indicating improvement or something positive, such as 'good news' (line 17), 'got better' (line 18), 'a good sign' (line 22), etc. Although there was a negative expression 'disappointed' in line 20, it directly followed by the conjunction 'but', indicating a shift from negative to positive in meaning. These positive expressions were important to understand the meaning of *LOOK up* in the collocation 'things + *LOOK up*'.

The co-occurrence of *LOOK up* with those positive expressions was probably restricted by a conceptual metaphor "GOOD is UP" proposed by Lakoff and Johnson (1980, p.16).

Lakoff and Johnson (1980) explain that this metaphor connects the physical basis of *up* with personal or social well-being. Corpus evidence of 'things + *LOOK up*' reflected that this collocation had a strong association with something good and positive (e.g. lines 17 and 22). Therefore, it can be argued that the collocation 'things + *LOOK up*' was associated with a metaphorical meaning of *LOOK up*, and this metaphorical meaning was labelled as *LOOK up*$_2$ in the thesis. *LOOK up*$_2$ was concerned with the improvement or a positive change of the situation because the non-verbal component *up* connoted a positive meaning in 'things + *LOOK up*'. Moreover, when *up* connoted a metaphorical meaning, the meaning of the verbal component became weak, losing the literal meaning of using eyes. In other words, the meaning of *LOOK up*$_2$ was partially compositional, with the meaning of *LOOK* being difficult to isolate. According to the second semantic criterion mentioned in Section 5.1.2, *LOOK up*$_2$ should be classified as a PV because the non-verbal component was used in its extended meaning and the verbal component lost its meaning in the context.

It is also interesting that the verbal component was frequently used with the progressive form (e.g. 'things are *looking up*') or expressions suggesting the commencement of the positive change (e.g. 'things begin to *look up*' in line 24 and 'things started to *look up*' in line 25). I will return to this specific co-occurrence in Section 6.2 and discuss the relationship between the progressive form and the meaning of this PV.

The examination of *LOOK up* in the contiguous form showed two distinctive meanings of this combination, one of which referred to using or raising eyes physically and the other was about the improvement or positive change of a situation. However, these two meanings failed to interpret certain instances where the contiguous form of *LOOK up*

appeared. For example,

26	He was	**looked up** to by his peers as a natural leader.
27	a guy I really admired and	**looked up** to, so even for people to
28	meant the world to me. I	**looked up** to him with pride because he
29	his young team is	**looking up** to him. Like it or not, he is the role model
30	we used to	**look up** to the head of department

In lines 26 to 30, *LOOK up*$_1$ and *LOOK up*$_2$ did not make sense in the context. The combination of *LOOK* and *up* seemed to have a new pattern in the five examples. In the five examples, *LOOK up* was directly followed by a preposition, *to*. Compared with *LOOK up*$_1$ and *LOOK up*$_2$ associated with specific collocational patterns, it seemed challenging to observe frequent nominal collocates that co-occurred with *LOOK up to*. However, when I looked into the context on both sides, the observation showed that *LOOK up to* often co-occurred with individuals who were admirable, such as 'a natural leader' (line 26) and 'a role model' (line 29), or those who were experienced and highly respected, such as 'the head of department' (line 30). As a result, in lines 26 to 30, the meaning of *LOOK up* was associated with having admiration or respect for someone. This new meaning, labelled as *LOOK up*$_3$, cannot make sense by simply combining *LOOK* and *up*; it required a third item, *to*, to follow the combination.

The use of *LOOK up*$_3$ should be classified as a PV because it cannot be predicted by the literal meanings of *LOOK* and *up*. Similar to *LOOK up*$_2$, the non-verbal component also connoted a metaphorical meaning in *LOOK up*$_3$, although the relevant metaphor in this case was different. In the case of *LOOK up*$_3$, the relevant metaphor was likely to be "HIGH STATUS is UP" (Lakoff and Johnson, 1980, p.16). This conceptual metaphor reflects the way people conceptualize social status and power, connecting higher physical position (i.e. *up*) with higher social status and power (Lakoff and Johnson, 1980). This metaphor helped to understand the meaning of *LOOK up*$_3$. Since *up* was used in a metaphorical

way in *LOOK up*$_3$, the combination should be classified as a PV based on the second semantic criterion.

6.1.2 *Look + up*: collocation and meanings.

When *up* is not adjacent to *LOOK*, the combination has a non-contiguous form. As mentioned in the methodology, when I examined the non-contiguous form of verb + particle combinations, the length of the intervening item between the verb and the particle did not exceed two words (Section 4.3.1). Moreover, since the focus of this section was the non-contiguous form of *LOOK up*, I used the query '{look/V} +* up' in CQPweb, which successfully retrieved all the instances in which the two components were separated by one or two words (Section 4.3.1).

Table 6.2 lists five nominal collocates that frequently appeared on the right side of the target combination. As Table 6.2 shows, it seems that *dictionary, Internet,* and *book*, have some common semantic properties in their meanings. The three collocates can be seen as sources of information. For example, dictionaries could be a useful source where people can consult the definitions of words, and the Internet could be the source for online news and information.

Table 6.2 Significant nominal collocates on the right side of *LOOK* +* *up* in the BoE corpus: frequency and *t*-score

Node	Collocates	Frequency	*t*-score
LOOK* +* *up	dictionary	42	6.47
	internet	31	5.44
	phone	14	3.44
	book	14	3.19
	sky	8	2.69

31 unsure of the meaning of a word, **look** it **up** in a **dictionary** and try using it
32 this was the time when I'd had to **look** the word **up** in a **dictionary**, after a reading
33 spelling rules and, more often, **looking** words **up** in the **dictionary.**
34 who didn't have to **look** anything **up** in the **dictionary**
35 and decided to **look** it **up** on the **Internet** and thought
36 ever heard of the disease so they **looked** it **up** on the **internet.**
37 the branch offices were **looking** them **up** on the **internet** then printing
38 pointed to numbers and the writer **looked** them **up** in a **book** that listed necessary
39 central reference library I **looked** it **up** in a **book** on metallurgy.
40 takes pictures of them and **looks** things **up** in his **books.**

In addition to the frequent collocates around *LOOK* + *up*, I also examined the patterns associated with the node when it co-occurred with *dictionary*, *internet*, and *book*. Lines 31 to 40 were extracted from the BoE corpus and demonstrated patterns associated with *LOOK* + *up*. When the node co-occurred with *dictionary* and *book*, the pattern can be written as '*LOOK* + *up* + in + dictionary/book'. When it co-occurred with *internet*, the pattern was '*LOOK* + *up* + on + internet'. As the two patterns show, the non-contiguous combination was followed by prepositions *in* or *on*, connecting the object that was being looked up with the source of information (e.g. '*look* the word *up* in a dictionary' in line 32). The examination suggested that the meaning of *LOOK* + *up* was related to finding a piece of information from sources such as a dictionary or the Internet. This meaning was distinct from the other three meanings discussed in Section 6.1.1 and it was therefore labelled as *LOOK up*$_4$. Meanwhile, the literal meanings of *LOOK* and *up* made little contribution to the interpretation of *LOOK up*$_4$, indicating that the meaning of *LOOK up*$_4$ was non-compositional. Given the first semantic criterion, the non-contiguous combination of *LOOK* and *up* should be classified as a PV when *LOOK up*$_4$ made sense in the context.

Another important point related to *LOOK up*$_4$ was that it allowed the movement of the non-verbal component. Here are some examples showing the flexible movement of *up*.

41 You'll need to **look up** a photo of this, as it's one of
42 grandmother that week and found her **looking up** a word in the dictionary
43 Annie had gone to the library to **look up** some materials for one of her classes
44 While my food was being prepared, I **looked up** the translation of what I'd just ordered
45 library or use your computer **look up** what it is that this drug actually does

Lines 41 to 45 provide corpus evidence to show the contiguous use of *LOOK up*$_4$. When the particle followed the verb directly, *LOOK up*$_4$ still made sense in these contexts. However, it seemed that the pattern associated with *LOOK up*$_4$ became less evident when the two components were positioned to be adjacent to each other.

When the non-contiguous combinations of *LOOK up* were retrieved in the BoE corpus, there were a number of concordance lines showing that personal pronouns such as *her*, *him*, *me*, *you*, etc. frequently appeared between *LOOK* and *up.* For instance,

46 he would go and **look** her **up**, see what happened
47 keep on with my acting, I can **look** her **up** in London, and she'll try to help
48 waiting for someone to **look** him **up** for two years – since he had phoned
49 I'll try to **look** him **up**, ask him a few questions
50 say we would come and **look** you **up** in England one day.

In contrast to lines 31 to 40 where *LOOK up*$_4$ frequently co-occurred with nouns such as *dictionary* and *book*, the non-contiguous combination of *LOOK* and *up* was separated by personal pronouns in lines 46 to 50. Meanwhile, the use of '*LOOK* + personal pronoun + *up*' in lines 46 to 50 was also different from the way that *LOOK up*$_3$ was used (see lines 26 to 30), even though *LOOK up*$_3$ co-occurred with personal pronouns (e.g. 'I *look up* to him with pride'). The examination of lines 46 to 50 indicated a connection between '*LOOK* + personal pronoun + *up*' and the action of visiting someone, and the label *LOOK up*$_5$ was used to mark this new meaning. In addition, '*LOOK* + personal pronoun + *up*' seemed to have a semantic preference of modal verbs because it often co-occurred with modal verbs such as *would*, *can*, and *will* (see lines 46, 47, 49 and 50). In line 46, for example, the speaker used 'would go and *look* her *up*' to express his intention to visit

'her'. Similarly, the phrase 'would come and *look* you *up* one day' in line 50 conveyed the possibility of visiting at some point in the future. As a result, '*LOOK* + personal pronoun + *up*' showed a strong tendency to co-occur with modal verbs indicating intention or possibility. Given the meaning of *LOOK up*$_5$, this tendency is not particularly surprising due to Sinclair's idiom principle, which suggests the co-selection of lexical items in the context.

According to the above analysis, it was evident that *LOOK up*$_5$ was associated with the action of visiting. Unlike *LOOK up*$_4$, this meaning was partially compositional because it was possible to connect the meaning of *LOOK* with the action of visiting or meeting someone. However, the meaning of *up* became weak when *LOOK up*$_5$ made sense in the context. Based on the second semantic criterion, *LOOK up*$_5$ was classified as a PV in the current study because the verbal component was used in an extended way and the non-verbal component was delexicalized in this combination.

Corpus evidence showed that *LOOK up*$_5$ also accepted the movement of the non-verbal component. However, the contiguous form of *LOOK up*$_5$ was extremely rare in the BoE corpus. Here are some examples showing the contiguous form of *LOOK up*$_5$ in the BoE corpus.

51	While in New York, Bing	**looked up** his old friend Fritz Stiedry,
52	and at her suggestion had	**looked up** Rosalind when they came to London.
53	a visitor might want to	**look up** her first cousin Bob Hancock,

In addition to *LOOK up*$_4$ and *LOOK up*$_5$, I also observed another frequent usage of *LOOK* + *up*, as demonstrated by the following concordance lines.

54 He stop, **looks** her **up** and down critically,
55 Ella remarked coolly, **looking** her **up** and down with a critical eye.
56 opened the door and **looked** him **up** and down suspiciously.
57 The beggar with the beard **looked** Laurence **up** and down, as if he were an intruder
58 He **looked** me **up** and down in disbelief.
59 He sneered and slowly **looked** me **up** and down as if I were for sale.
60 you report two girls for **looking** you **up** and down and turning away with a sneer.

The five different meanings of *LOOK up* cannot interpret the meaning of *LOOK* + *up* in lines 54 to 60. The combination of *LOOK* and *up* seemed to be part of a longer sequence, '*LOOK* someone *up* and down'. A closer examination showed that the word choice between *LOOK* and *up* was highly restricted, only accepting personal pronouns such as 'her', 'him', 'Laurence', etc. The structure of this sequence was also fixed. There was no opportunity to change the order of the sequence (e.g. **look up her and down* or **look her down and up*).

I also examined the surrounding information in the co-text of "*LOOK* someone *up* and down'. An important observation was that this sequence frequently co-occurred with expressions indicating negative attitudes, such as 'with a critical eye' (line 55), 'suspiciously' (line 56), 'in disbelief' (line 58). The co-occurrence of '*LOOK* someone *up* and down' with these expressions suggested a negative semantic prosody of this sequence, connecting the action of looking someone up and down with a harsh and critical judgment. Also, L1 users employed expressions indicating impolite behaviours in the co-text of this sequence, such as 'sneered' (line 59) and 'turning away with a sneer' (line 60). As a result, the examination of '*LOOK* someone *up* and down' indicated that this phrase was associated with judging someone in a disrespectful and impolite manner.

6.1.3 Meanings and patterns associated with *LOOK up* in the BoE corpus

After examining the collocation of *LOOK up* in the BoE corpus, I have identified five distinctive meanings of this combination and one fixed phrase, '*LOOK* someone up and down'. Corpus evidence of *LOOK up* supported the hypothesis that collocation is an important indicator of specific meanings of PVs. Table 6.3 summarises meanings and frequent patterns associated with *LOOK up* when it co-occurred with different collocates.

Table 6.3 An overview of different meanings and patterns associated with look up in the BoE corpus.

	Meanings	**Patterns**
***LOOK up*$_1$**	to use or raise eyes physically in order to see someone or something	- *LOOK up* into one's eyes - *LOOK up* with + mood-associated adjective + eyes - *LOOK up* at the sky/ceiling/star
***LOOK up*$_2$**	to improve or become better	- things + *LOOK up* (usually used with the progressive form)
***LOOK up*$_3$**	to admire or respect someone	- *LOOK up to* someone
***LOOK up*$_4$**	to find a piece of information from a source, such as a dictionary or the Internet	- *LOOK* + *up* + in + dictionary/book - *LOOK* + *up* + on + internet
***LOOK up*$_5$**	to visit somebody	- *LOOK* + personal pronoun + *up*

Different meanings of *LOOK up* also provide evidence to support that verb + particle combinations usually have various usages. Table 6.4 below shows the classification of *LOOK up* when it was used in different meanings.

Table 6.4 Classifying the use of *LOOK up* in the BoE corpus

	Meanings	**Compositionality**	**Classification**
***LOOK up*$_1$**	to use or raise eyes physically to see someone or something	Compositional	Non-PV
***LOOK up*$_2$**	to improve or become better	Partially compositional	PV
***LOOK up*$_3$**	to admire or respect someone	Partially compositional	PV
***LOOK up*$_4$**	to find a piece of information from a source, such as a dictionary or the Internet	Non-compositional	PV
***LOOK up*$_5$**	to visit somebody	Partially compositional	PV

In the previous chapter, I have argued that the classification of verb + particle combinations, either as PVs or non-PV combinations, should be determined by the context in which it appears (see Section 5.2). The analysis of *LOOK up* provided corpus evidence to support this argument and highlighted the important role of collocation in understanding the meanings of the verb + particle combination. The concordance lines showed that the combination of *LOOK* and *up* can be used either as a non-PV combination or a PV. When *LOOK up* co-occurred with nouns such as *eye, sky, ceiling,* it was used as a non-PV combination because these nouns were associated with the literal meanings of *LOOK* and *up* (i.e. *LOOK up*$_1$). However, when the combination co-occurred with nouns such as *book, dictionary,* and *internet*, it should be classified as a PV because it was hard to isolate the meaning of each component in the combination (i.e. *LOOK up*$_4$). The findings also showed that the non-verbal component *up* sometimes connoted metaphorical meanings in the combination (i.e. *LOOK up*$_2$ and *LOOK up*$_3$). For example, the conceptual metaphor 'GOOD is UP' helps to understand *LOOK up*$_2$. Corpus evidence showed that expressions associated with positive evaluation, such as 'good news' and 'a good sign', often appeared in the co-text of the collocation 'things + *LOOK*

up'. The semantic preference of *LOOK up*$_2$ also supported the argument that "upward orientation tends to go together with positive evaluation" (Kövecses, 2010, p.40).

Finally, the current study also investigated the most frequent use of *LOOK up* in the BoE corpus. Table 6.5 shows the frequency of different usages of this combination in the 500 random samples.

Table 6.5 Frequency of the use of *LOOK up* in 500 random concordance lines

	Use	**Frequency**	**Percentage (%)**
***LOOK up*$_1$**	Non-PV	355	71
***LOOK up*$_2$**	PV	21	4.2
***LOOK up*$_3$**		17	3.4
***LOOK up*$_4$**		64	12.8
***LOOK up*$_5$**		7	1.4
***LOOK* someone *up* and down**	Fixed phrase	36	7.2

As Table 6.5 shows, the non-PV use, *LOOK up*$_1$, was more frequent than other usages in the BoE corpus. Among the 500 random concordance lines, the PV usages appeared 128 times and the fixed sequence appeared 17 times. This finding is consistent with Garnier and Schmitt's (2015) finding. Garnier and Schmitt (2015) regarded the literal meaning of *LOOK up* (i.e. *LOOK up*$_1$) as the most important meaning of this combination because of its high frequency. Although *LOOK up* has more meanings when it was used as a PV, the frequency of PV usages was less frequent in the BoE corpus. Moreover, the frequency of the fixed phrase, *LOOK* someone *up* and down, was even higher than some PV usages (see Table 6.5), indicating the significance of this special phrase.

6.2 The impact of the progressive form on the meanings of PVs

6.2.1 *LOOK up*$_2$: the progressive form of 'things + *LOOK up*' and its association with future predictions

In the previous section, concordance results of the collocation 'things + *LOOK up*' showed that *LOOK up*$_2$ tended to be used in the progressive form frequently (see lines 16 to 25). In the BoE corpus, there were three different forms of this collocation and Table 6.6 provides information about the frequency of each form.

Table 6.6 Frequency of 'things + *LOOK up*' in the BoE corpus

Forms	No. of occurrences	Percentage (%)
things + *looking up*	314	80.51
things + *look up*	55	14.1
things + *looked up*	21	5.38
Total	390	100

As Table 6.6 shows, the collocation 'things + *LOOK up*' was predominantly used in its progressive form (i.e. 'things + *looking up*') in the BoE corpus. As mentioned in Section 3.3, the progressive form is usually associated with three major meanings: temporariness, duration, and habit (Quirk et al., 1985; Mindt, 1997, 2000). However, in the concordance lines where 'things + *looking up*' appeared, there was no evidence showing the association between the progressive form and the three major meanings. For example,

> Good news is that things are ***looking up***. He will be back. (BoE: NI2__990599)
>
> Mark Hughes, the manager, believes that things are ***looking up*** for the club. (BoE: NB1__041120)

A closer examination of the two examples showed that L1 users intended to predict the future situation via the use of 'things + *looking up*'. In the first example, the contextual

information ‘he will be back’ suggested that ‘things are *looking up*’ described a prediction about the future. It was also evident in the second example that ‘things are *looking up* for the club’ was a prediction rather than a description of the current situation. As a result, the progressive form of the collocation ‘things + *LOOK up*’ has the function to predict the future rather than describe the current situation. This finding provides corpus evidence to support Mindt’s (1997) argument about the prediction meaning of the progressive form.

In addition to the progressive form, Table 6.6 also shows L1 users employed the simple form (i.e. things + *look up*) and the past form (i.e. things + *looked up*) sometimes. Here are some examples illustrating the use of these two less frequent forms:

61 confident that **things** will **look up** once the company broadens its product portfolio
62 and **things** started to **look up**, then Newcastle were supposed to be interested
63 **Things** begin to **look up** when Danny starts a band and a beautiful blonde

64 **Things** **looked up** in 1987, although much of the $25 million
65 Although **things** have **looked up** for British Airways since last quarter's ghastly
66 **Things** have certainly **looked up** from what they were 17 months back when Gogoi

As lines 61 to 63 show, the simple form ‘things + *look up*’ also implied the future situation because it was part of the future tense. For instance, ‘things will *look up*’ in line 61 was a typical use of the future tense. Moreover, expressions, such as ‘started to’ (line 62) and ‘begin to’ (line 63), could imply a future domain on the basis of the simple form although they were not typical future adverbials[25] in English. Declerck, Reed & Cappelle (2006) use the label ‘futurish form’ to refer to expressions that “establish a future domain but are not future tense forms” (p.106). The use of the future tense and futurish forms connected the simple form, ‘things + *look up*’, with a situation that will

[25] Future adverbials are expressions that often co-occur with verbs in future tenses, specifying when and how an event will happen in the future (Quirk et al., 1985). For example, *next week* is a common future adverbial indicating when an event will happen.

happen in the future. Therefore, when 'things + *look up*' appeared, the meaning was also associated with the future prediction. Corpus evidence showed that this association can be triggered by specific futurish forms, such as 'things will *look up*', 'things started to *look up*', 'things begin to *look up*', etc.

Lines 64 to 66 demonstrate the use of the past form 'things + *looked up*'. Unlike 'things + *looking up*' and 'things + *look up*', there was no clue showing the association between 'things + *looked up*' and the future prediction. In lines 65 and 66, the past form co-occurred with the present perfect tense, as in 'things have *looked up* for British Airways' and 'things have certainly *looked up*'. Declerck, Reed & Cappelle (2006) argue that using the present perfect can imply the influence of a past event on the present. While there were adverbials indicating when the events took place in the past (e.g. 'since last quarter' in line 65 and 'from what they were 17 months back' in line 66), L1 users employed the present perfect to emphasise the influence on the present rather than the past. As a result, 'things have *looked up*' indicated an improved situation in the present, signifying the realisation of a prediction made in the past.

6.2.2 *LOOK for*: meanings and the association with the present progressive form.

In addition to *LOOK up*$_2$, the corpus data showed that the progressive form frequently co-occurred with other verb + particle combinations, such as *LOOK for* and *LOOK forward to*. This section focuses on *LOOK for* and Section 6.2.3 deals with *LOOK forward to*. In the BoE corpus, there are 67678 concordance results containing *LOOK for* and four forms of this PV can be found (i.e. *look for, looks for, looked for* and *looking for*). Table 6.7 shows the frequency of each form.

Table 6.7 Frequency of *LOOK for* in the BoE corpus

Forms	Frequency	Percentage (%)
looking for	49225	72.73
look for	14882	21.99
looked for	2360	3.49
looks for	1211	1.79
Total	67678	100

The frequency breakdown in Table 6.7 shows that the progressive form, *looking for*, was the most frequent form in the BoE corpus. The progressive form of *LOOK for*, found in 49,225 concordance lines, accounted for 72.73% of the total occurrences. The high proportion suggested that L1 users preferred the progressive form over the other three forms. When *LOOK for* was used in the progressive form, corpus evidence showed that the present progressive was more frequent, occurring 14586 times, whereas the past progressive only appeared in 6320 concordance lines. In Section 6.2.1, the analysis of 'things + *looking up*' showed that the communicative function of the progressive form was to make predictions. In terms of *looking for*, I also analysed the concordance lines in which it was used in order to investigate the communicative function.

When L1 users used *LOOK for* in the progressive form, they tended to describe actions that are in progress and has not yet been completed. For example,

67 and now they are **looking for** evidence to back it up.
68 cops right now are probably **looking for** a big black guy who dragged
69 bachelor degree graduates **looking for** full-time work this year,
70 And while you are **looking for** a safe place to move into, you need someone who
71 When you are **looking for** a job, do not forget that under the Sex
72 They are not **looking for** a solution to a problem as they tell their tales.

As lines 67 to 72 show, there were two types of time adverbials indicating the duration of *looking for*. The first type was to use explicit adverbials, such as 'now', 'right now', 'this year' (lines 67 to 69), to indicate the time when *looking for* is taking place. In lines 70 to 72, L1 users indicated the time in an implicit way. For example, *looking for*

appeared in the *while*-clause in line 70, which suggested the action of '*looking for* a safe place' is in progress. Also, *looking for* co-occurred with the *as*-clause in line 72, indicating that the action in the main clause (i.e. 'not *looking for* a solution') and that in the *as*-clause (i.e. 'tell their tales') are taking place simultaneously. When *looking for* co-occurred with explicit and implicit time adverbials, the meaning tended to be associated with searching for somebody (e.g. '*looking for* a big black guy' in line 68) or something (e.g. '*looking for* a safe place' in line 70). This meaning was labelled as *LOOK for*$_1$ in the current study. The examination of *LOOK for*$_1$ showed that the progressive form of *LOOK for*$_1$ often co-occurred with time adverbials such as 'now', 'this year', *while*-clause, etc., implying the duration of searching. However, it should be noted that the duration meaning associated with *LOOK for*$_1$ was influenced by the progressive form rather than by the co-occurrence of time adverbials in the co-text. The following corpus examples can support this argument.

73 The county is **looking for** a new contractor to finish the job.
74 These people are **looking for** work or training in Japan.
75 Some NHS organisations are **looking for** IT workers with sound knowledge of IT

As lines 73 to 75 show, while there was no time adverbial in the co-text, the progressive form of *LOOK for* implied that the action of searching is in progress and has not yet finished.

Besides *LOOK for*$_1$, corpus evidence showed that the progressive form of *LOOK for* frequently followed an adverb *always* (i.e. 'always *looking for*'). In the BoE corpus, there were 686 instances showing the occurrence of *always* at the N-1 position around *looking for*. Lines 76 to 80 are examples of 'always *looking for*'.

76	He is always	**looking for**	ways to obtain money, but spends it like water
77	We're always	**looking for**	them, but they're harder to trap than rats
78	You're always	**looking for**	a job, even if you already have a job
79	Children are always	**looking for**	new adventures to try but anyone who plays
80	I'm always	**looking for**	positive things, but my feeling is that

The examination of lines 76 to 80 showed that 'always *looking for*' did not indicate the duration of the action, which was different from *LOOK for*$_1$. *Looking for* became a habitual action in these contexts because of the use of *always*, an adverb describing the frequency of an action. When the action of *looking for* repeated habitually, L1 users tended to convey an emotional meaning that exceeds 'searching'. The emotional meaning of *looking for* can be interpreted as 'expecting or longing for', and this emotional meaning was labelled as *LOOK for*$_2$ in the current study. Moreover, it seemed that L1 users showed their complaints or disagreement via the use of 'always *looking for*'. For example, in line 76, the speaker expressed a complaint about the subject who spent money like water. In line 78, the communicative meaning of the sentence suggested the speaker's disagreement. The findings of 'always *looking for*' supported the argument that the progressive form can indicate speakers' negative attitude when it co-occurs with *always* (Römer, 2005; Lindley, 2020). Moreover, Lindley (2020) also showed that when the progressive form was used to convey a negative meaning, it was less likely to observed negative words in the co-text. With regard to 'always *looking for*', there was no negative word in the co-text, but L1 users tended to use 'but' or 'even if' to suggest a soft negation (see lines 76 and 78). The negative meaning of 'always *looking for*' seemed to be clearer due to the soft negation.

Corpus evidence also showed that the progressive form of *LOOK for* frequently followed another adverb, *really*. In the BoE corpus, there were 176 concordance lines showing the occurrence of *really* at the N-1 position around *looking for*. However, unlike 'always

looking for' that conveyed complaints or disagreement, it seemed that 'really *looking for*' only connoted an emotional meaning. The combination of *really* and *looking for* did not have a negative or positive connotation. For instance,

81 What I'm really **looking for** is purpose and enjoyment each day
82 clear that he is really **looking for** an answer. If he wants an answer
83 All I'm really **looking for** here is a good reason to take your name off my list
84 If you're really **looking for** a bargain, request space in a lodge
85 We're really **looking for** a partner and a friend to be with us for years

The examination of lines 81 to 85 showed that L1 users employed *really* to modify the emotional meaning of *LOOK for* (i.e. *LOOK for*$_2$). For instance, the use of *really* in line 82 suggested the longing for 'an answer'. The emotion seemed to be stronger with the help of *really*. In addition to *always* and *really*, corpus evidence showed that *LOOK for*$_2$ also co-occurred with other adverbs, such as *certainly*, *definitely*, and *obviously* (see lines 86 to 88). However, L1 users also expressed a neutral emotional meaning in these contexts.

86 we are certainly **looking for** a sustained improvement in weather conditions
87 We're definitely **looking for** a chance to get a little revenge
88 Employers are obviously **looking for** the best qualified to fill a vacancy

Based on the analysis of *LOOK for* in the progressive form, I have identified two different meanings of this combination. *LOOK for*$_1$ was associated with the action of searching and it often co-occurred with explicit or implicit time adverbials such as 'now', 'this year', *as*-clause, etc. *LOOK for*$_2$, on the other hand, frequently appeared in the context where no time adverbials can be observed. Corpus evidence showed that *LOOK for*$_2$ was associated with a greater level of emotion and it often followed adverbs such as *always, really,* etc. When the progressive form of *LOOK for* was combined with *always*, L1 users tended to connote a negative meaning in the context (see lines 76 to 80), whereas the emotional meaning tended to be neutral when *looking for* co-occurred with other

adverbs (see lines 81 to 88). Table 6.8 lists the two different meanings of *LOOK for* and provides examples for each meaning.

Table 6.8 Meanings and examples of *LOOK for* in the BoE corpus

	Meanings	Examples
***LOOK for*$_1$**	To search for	They are not ***looking for*** a solution to a problem as they tell their tales. (line 72)
***LOOK for*$_2$**	To expect or long for	He is always ***looking for*** ways to obtain money, but spends it like water. (line 76) What I'm really ***looking for*** is purpose and enjoyment each day. (line 81)

6.2.3 *LOOK forward to*: differences between the present progressive and the past progressive

In terms of *LOOK forward to*, corpus data also showed the progressive form, *looking forward to*, was more frequent than other forms. As Table 6.9 shows, among the 19133 occurrences of *LOOK forward to* in the BoE corpus, the progressive form appeared in more than half of the total occurrences (55.09%). L1 users were less likely to choose the past form (*looked forward to*) and the third-person singular form (*looks forward to*) due to their relatively low frequency in the BoE corpus, accounting for only 6.49% and 4.22% respectively.

Table 6.9 Frequency of *LOOK forward to* in the BoE corpus

Forms	Frequency	Percentage (%)
looking forward to	10540	55.09
look forward to	6544	34.20
looked forward to	1241	6.49
looks forward to	808	4.22
Total	19133	100

Lines 89 to 93 were chosen to demonstrate the progressive form of *LOOK forward to*. The examination showed that the meaning of *looking forward to* was associated with

anticipation or expectation. For example, in line 90, the speaker expected the live music because the speaker has not ‘heard any live music for months’.

89 I'm **looking forward to** meeting her.
90 I'm **looking forward to** this. I haven't heard any live music for months.
91 She is **looking forward to** seeing you again and is planning a little party for
92 Michael is **looking forward to** his day in court and wishes to thank the millions of fans
93 we are **looking forward to** a variety of interesting talks and other events.

Moreover, when *LOOK forward to* was used in the progressive form, L1 users did not employ explicit or implicit time adverbials in the co-text. L1 users tended to use *looking forward to* to convey a tentative meaning. The progressive form can add a sense of tentativeness to the utterance, making it less direct and more polite (Quirk et al., 1985; Mindt, 1997; 2000). The tentative meaning can be illustrated by the five concordance examples. For instance, in line 91, the communicative function of *looking forward to* was to express the speaker’s tentativeness and uncertainty about the timing of ‘seeing you again’.

Lines 89 to 93 showed the tentative use of *LOOK forward to* in the present progressive form. In the BoE corpus, L1 users also used the past progressive of this PV frequently (i.e. was/were *looking forward to*). Here are some concordance lines showing the past progressive form of *LOOK forward to*.

94 I had trained for months and was really **looking forward to** it but then I got sick.
95 I was almost **looking forward to** moving on. But now he's come back into my life and I'm not sure about anything anymore.
96 I was **looking forward to** something really passionate, but my boyfriend was so quick and I didn't feel much.
97 Mike said he was **looking forward to** the music but said it was getting too expensive and straying from its country roots.
98 I was **looking forward to** playing in both events but in the present climate I do not feel comfortable flying.

The examination of *looking forward to* in lines 94 to 98 showed that L1 users tended to express a sense of regret or disappointment by using the past progressive. For instance, in line 97, ‘Mike’ felt disappointed with ‘the music’ due to the expensive price and the

change in style. Also, in line 98, 'playing in both events' was not going to happen due to the uncomfortable climate. Based on the analysis, the past progressive form of *LOOK forward to* tended to suggest the speaker's disappointment, which was different from the way of using the present progressive form. In addition, corpus evidence showed that L1 users did not combine *looking forward to* with *always* frequently. There were only 12 concordance lines in the BoE corpus containing the occurrence of 'always *looking forward to*'. When *always* occurred in the co-text, *looking forward to* was used in the past tense. For example,

99 She was always **looking forward to** the future, anxious to plan and develop exciting new things for Atlantis.

100 He says that they were always **looking forward to** doing their next movie together and never got the chance.

The communicative meanings of *looking forward to* in the two examples were different. The use of 'always *looking forward to*' connoted a positive expectation in line 99, whereas it was associated with a negative feeling of disappointment in line 100.

Based on the examination of *looking forward to* in the present tense and the past tense, it can be argued that the use of different tense forms led to different communicative meanings. L1 users tended to express a tentative meaning via the use of the present progressive form (see lines 89 to 93). However, when 'was/were *looking forward to*' occurred, the past progressive form was associated with regret and disappointment (see lines 94 to 98).

6.3 Summary

To sum up, this chapter has shown the collocation and lexical patterns associated with a verb + particle combination, *LOOK up*. Based on the collocates in the co-text of *LOOK*

up, the current study has identified five distinctive meanings of this combination. In other words, collocates in the co-text help to differentiate the meanings of polysemous PVs. For example, *LOOK up*$_1$ co-occurred with nouns such as *eye* and *sky*, whereas the two collocates cannot occur in the co-text of *LOOK up*$_4$. In the BoE corpus, the frequent collocates of *LOOK up*$_4$ were *dictionary, book* and *internet*. The findings support Sinclair's idiom principle in that the choice of collocates in the co-text is restricted by the meaning of the node (Sinclair, 1991). Moreover, the examination also showed that different forms of *LOOK up* were associated with different meanings. For example, when this PV was used in the non-contiguous form, the two metaphorical meanings (i.e. *LOOK up*$_2$ and *LOOK up*$_3$) did not make sense, as they were specifically associated with the contiguous form.

In terms of the progressive form, Section 6.2 showed how this tense form influenced the communicative meanings of three PVs, *LOOK up*$_2$, *LOOK for*, and *LOOK forward to*. Corpus data of 'thing + *looking up*' suggested that the function of *LOOK up*$_2$ was associated with making a positive predication about the future. In terms of *LOOK for*, when this PV co-occurred with time adverbials such as 'now', 'this year' and *as*-clause, it usually conveyed a duration meaning. When *LOOK for* was combined with adverbs such as *always* and *really*, it suggested an emotional meaning. However, the combination 'always *looking for*' had a special use, suggesting complaints or disagreement. Finally, the examination *of LOOK forward to* showed that the present progressive and the past progressive were used to express different communicative meanings. The present progressive frequently conveyed a tentative meaning, whereas the past progressive was associated with a negative feeling of regret or disappointment.

Chapter 7 Grammatical patterns associated with PVs

In the previous chapter, I have discussed the lexical patterns associated with a specific PV based on the investigation into collocation in the co-text. This chapter shifts the focus to grammatical patterns associated with specific PVs. It also discusses the challenges of replacing PVs with one-word verbs from the perspective of pattern grammar. As mentioned previously in Section 2.2.2, a traditional approach to identifying PVs is based on the idea that some PVs can be replaced with one-word verbs (Bolinger, 1971; Biber et al., 1999). For example,

> The public house next to the cinema had just ***exploded***. Another bomb in another nearby pub ***blew up*** a few minutes later. (BoE: BB_Em012020)

It is possible for *exploded* to replace the PV, *blew up*, because both of them are associated with 'an explosion' in the given example. However, while *explode* and *BLOW up* can be used interchangeably in the given example, the polysemous feature of PVs suggests that *BLOW up* may have a wider range of meanings that cannot be replaced with *explode*. Here is a corpus example in which '*blew up* at her' fails to be substituted by '**explode* at her'.

> She just doesn't understand why you ***blew up*** at her like that. (BoE: BUCXF002456)
>
> *She just doesn't understand why you ***exploded*** at her like that.

The replacement is unsuccessful because the PV is used in a figurative way that is not synonymous with *explode*. In this example, the meaning of '*blew up* at her' seems to be associated with 'losing one's temper' (OED online). However, '**exploded* at her' does not convey the same meaning.

The example of '*blew up* at her' provides evidence to show the difficulty of replacing PVs with one-word verbs. Previous studies have also noted the risk of such replacements by demonstrating the stylistic differences between PVs and one-word verbs (e.g. Side, 1990; Siyanova and Schmitt, 2007). Side (1990) suggested that *GIVE out* and *distribute* tended to be used in different registers. The individual verb was likely to be used in written and formal register (Side, 1990). However, in the following example, it is unlikely to replace *giving out* with *distributing* because 'hate *giving out* leaflets' conveys a sense of complaint which the equivalent *distributing* does not.

> Salesmen hate ***giving out*** leaflets to people who are only interested in getting a freebie. (BoE: MB4__050307)

Siyanova and Schmitt (2007) pointed out that PVs tended to have a colloquial tone and usually appear in spoken and informal register, whereas one-word verbs are more likely to appear in written discourse such as academic writing or official reports. The stylistic difference between PVs and one-word verbs suggests that "direct equivalents of phrasal verbs do not always exist" (Side, 1990, p.145).

This chapter aims to show the differences between PVs and one-word verbs from the perspective of pattern grammar. Section 7.1 outlines the differences between *LOOK for*$_1$ and *search*, arguing that *search* fails to replace the PV because of the distinctive grammatical patterns. Section 7.2 focuses on *LOOK up* and examines the semantic nuances when *LOOK up to* and its two potential alternatives, *admire* and *respect*, appear in the same pattern. Finally, Section 7.3 discusses situations in which PVs cannot be replaced by one-word verbs.

7.1 The grammatical differences between PVs and one-word verbs

As mentioned in Section 2.3.1, many studies have shown that L2 learners frequently avoid using multi-word verbs in both written and spoken discourse (Dagut and Laufer, 1985; Hulstijn and Marchena, 1989; Liao and Fukuya, 2004; Schmitt and Siyanova, 2007). They tend to choose one-word verbs that has an equivalent meaning and regard it as a strategic approach to avoid using PVs (Hulstijn and Marchena, 1989). In Section 6.2, I have outlined two meanings of *LOOK for*, one of which is associated with searching (i.e. *LOOK for*$_1$). As a result, the verb *search* is possible to be a potential alternative for *LOOK for*$_1$ because their meanings are related. The study firstly investigated the patterns in which the individual verb *search* was frequently used. The following concordance lines show the use of *search* in the BoE corpus.

1		He decided to	**search** his bag for the scrap of paper with the phone number
2		he told them that he'd	**searched** the hotel room thoroughly, but that there was no sign
3		He	**searched** his pockets for a coin to put in.
4		but I've been out	**searching** the streets for him all night.
5		Squads of soldiers were	**searching** the flooded ruins in the hope of finding survivors
6		I had no opportunity to	**search** for answers to these questions until I arrived
7		slowed again as he	**searched** for adequate words to describe the events of that
8		she	**searched** for the special security key to enable her to open
9		Selby was just	**searching** for more paper and another pencil
10		I've heard you're	**searching** for gifted apprentices.

The 10 lines of *search* showed two distinctive usages of this verb. Lines 1 to 5 showed the transitive use of *search*, whereas lines 6 to 10 illustrated the intransitive use. As a transitive verb, corpus evidence suggested the frequent pattern in which *search* often appeared was **V n *for* n**. As lines 1 to 5 showed, nouns such as 'bag', 'pockets', 'hotel room', 'streets', 'flooded ruins' can appear in the first noun slot (i.e. between the verb and the preposition *for*). It seemed that the choice of these nouns was not random, as they were associated with the container or place in which the action of searching took place. The second noun slot in this pattern, such as 'the scrap of paper' (line 1) and 'a

coin' (line 3), indicated the aim of searching. As a result, '*search n for n*' can be understood as searching somewhere for something.

In addition to *search*, Francis, Hunston and Manning (1996) showed that the **V n *for* n** pattern was also applicable to a group of semantically associated verbs, including *comb, explore, scan, scour,* etc. Lines 11 to 15 below provide examples to illustrate the use of these verbs in the pattern **V n *for* n**. Compared with '*search n for n*', the examination of these five lines suggested that there was no significant change in the structure and meaning when *comb, explore, scan,* etc. appeared in the verb slot.

11		Sir, I've been	**combing** the files **for** anything about Cardinal.
12		Restless prospectors were	**exploring** the world **for** more gold
13		American forces	**scanned** the area **for** the number of dead bodies.
14		his other main hobby -	**scouring** the woods **for** new and unfamiliar species
15	me to go on ahead of them to		**scout** the area **for** any signs of the military.

In terms of *LOOK for*$_1$, however, there was no corpus evidence showing that this PV can be used in the pattern **V n *for* n**. In the BoE corpus, L1 users employed *LOOK for*$_1$ in a pattern that was different from *search*. The following concordance lines show the grammatical pattern associated with *LOOK for*$_1$ and demonstrate the differences between *LOOK for*$_1$ and *search.*

16		we're going to	**look for** a cottage in the Highlands.
17		Phil thinks we should	**look for** a place in Moseley Village.
18		I'd	**look for** him in the locker room.
19		then you had to	**look for** someone in a certain age-group
20	Omaxin and pretended to		**look for** the answers in the ruins.

As lines 16 to 20 show, the frequent pattern in which *LOOK for*$_1$ appeared was **V n *in* n**. In this pattern, the first noun indicated the aim of searching (e.g. 'the answers') whereas the second indicated the place where the action of searching took place (e.g. 'the ruins'). The meaning of '*LOOK for n in n*' should be interpreted as searching for something in somewhere. In other words, the syntactic order of nouns in '*LOOK for n in n*' was

different from the order in '*search n for n*'.

The study also tested whether the transitive verb, *search*, could appear in the **V n *in* n** pattern. Unfortunately, the unsuccessful attempts illustrated by lines 16* and 20* implied that the transitive *search* was incompatible with the **V n *in* n** pattern.

16* we're going to ***search** a cottage in the Highlands.
20* Omaxin and pretended to ***search** the answers in the ruins.

Based on the examination of *search* and *LOOK for*$_1$, it can be argued that the grammatical patterns in which *search* and *LOOK for*$_1$ appeared were different. Corpus evidence showed that the verb *search* was frequently used in **V n *for* n** and this pattern only allowed noun phrases indicating the place to follow the verb directly (e.g. 'the hotel room' and 'the street'). However, *search* cannot be used in the pattern where *LOOK for*$_1$ frequently appears (i.e. **V n *in* n**). In **V n *in* n**, it is the second noun that indicated where the action took place. Therefore, it is impossible to replace *LOOK for*$_1$ with the transitive *search*.

Moving on to the intransitive use of *search*, as lines 6 to 10 above show, **V *for* n** was the frequent pattern in which the intransitive *search* can be used. In addition to the intransitive *search*, Francis, Hunston and Manning (1996) observed that other verbs, such as *dig, explore, grope, hunt, scan*, were also used in this pattern. Meanwhile, they had some common semantic features with *search*. For example,

21 come all this way to **dig for** the treasure which, they say, is still buried
22 a license from the king to **explore for** metals in New Mexico.
23 she rolled over and **groped for** the alarm clock, peering at it
24 The companies will **hunt for** oil in several areas off Burma's coast
25 or smart cameras that **scan for** suspicious behaviour or things

Francis, Hunston and Manning (1996) also included the verb *look* in this 'search' group.

From the authors point of view, *search* and *look* can be seen as a pair of near synonyms because both can be used in the **V *for* n** pattern. However, it is impossible for the PV *LOOK for*$_1$ to appear in **V *for* n**. Since the intransitive *search* was also used in a different way from *LOOK for*$_1$, it can be argued that the replacement between *LOOK for*$_1$ and *search* cannot be successful, even though their meanings were closely related.

When I examined the **V *for* n** pattern, it was interesting to find that some PVs can be used in this pattern. Here are some KWIC lines extracted from the BoE corpus.

```
26          I've been told to look out for several landmarks on our approach to the airfield
27 wages were very low and he looked out for a more responsible, lucrative position in every
28                  I started looking out for product recalls.
29  exercises first. Learn to watch out for signs of tension during the day.
30     Mrs Jackson told me to watch out for a young man with very black hair.
```

Lines 26 to 30 show that *LOOK out* and *WATCH out* are two lexicalised PVs that are compatible with **V *for* n**. The examination of lines 26 to 30 suggested that the meanings of *LOOK out* and *WATCH out* were generally associated with searching, though there were some semantic nuances implied by *LOOK out* and *WATCH out*. For instance, when L1 users employed '*look out* for product recalls' (line 23) and '*watch out* for signs of tension' (line 24), they tended to warn someone to be cautious about something risky or unsafe (i.e. 'product recalls' and 'signs of tension'). Therefore, when *LOOK out* and *WATCH out* were used in the pattern **V *for* n**, the communicative meanings of the two PVs were associated with searching for something risky or unsafe. The individual verb *search* did not have this connotation. This semantic nuance between '*search for n*' and '*LOOK out/WATCH out for n*' suggested that *search* cannot be seen as a satisfactory alternative for *LOOK out* or *WATCH out*.

To sum up, the examination of grammatical patterns associated with *LOOK for*$_1$ and its

potential alternative *search* showed that it was not possible to replace *LOOK for*$_1$ with *search*. In the BoE corpus, *LOOK for*$_1$ frequently appeared in the pattern **V n *in* n**, which was different from the pattern in which *search* was used. Corpus evidence showed that *search* often appeared in **V n *for* n** and **V *for* n**. Due to grammatical differences between this pair of near synonyms, it can be argued that *LOOK for*$_1$ cannot be replaced with *search.* This finding supported Side's (1990) argument that "direct equivalents of phrasal verbs do not always exist" (p.145). Besides grammatical differences, the findings from '*search for n*' and '*LOOK out/WATCH out for n*' suggested that semantic nuances also led to an unsuccessful replacement.

7.2 The difficulty of finding satisfactory alternatives for polysemous PVs

In Section 7.1, the failure of replacing *LOOK for*$_1$ with *search* revealed that *search* was not a satisfactory alternative although its meaning was close to the PV. Prioritizing the semantic equivalence may result in the ignorance of grammatical differences between the PV and the potential one-word alternative. This section continues to investigate the replacement between PVs and one-word verbs but focuses on polysemous PVs. In this section, I chose the polysemous PV, *LOOK up*, as an example and examined if this polysemous PV can be replaced by one-word verbs successfully. In S ection 6.2, I have identified five distinctive meanings of *LOOK up* in the BoE corpus. For each meaning, Table 7.1 below lists some one- or multi-word verbs with the potential to replace the PV.

Table 7.1 Various meanings of *LOOK* up and potential one- or multi-word alternatives for each meaning

	Meanings	**Potential alternatives**
***LOOK up*$_1$**	to use or raise eyes physically in order to see someone or something	
***LOOK up*$_2$**	to improve or become better	*be/get better*
***LOOK up*$_3$**	to admire or respect someone	*admire* or *respect*
***LOOK up*$_4$**	to find a piece of information from a source, such as a dictionary or the Internet	*consult the dictionary*
***LOOK up*$_5$**	to visit somebody	*visit*

The primary criterion for finding alternatives in Table 7.1 was semantic equivalence. As Table 7.1 shows, it was not possible to find one replacement for a PV with multiple meanings. Different meanings of *LOOK up* seemed to be synonymous with different individual verbs or multi-word expressions. For example, *LOOK up*$_3$ might be replaced with *admire* or *respect* because these two verbs were synonymous with *LOOK up*$_3$; both the PV and the potential alternatives indicated the way of showing admiration or esteem towards a person. However, it was hard to find an appropriate replacement for the literal meaning, (i.e. *LOOK up*$_1$). In some cases, the PV did not have a one-word equivalent; the meaning tended to be associated with a multi-word expression (e.g. *LOOK up*$_2$ and *LOOK up*$_4$). More information regarding the multi-word replacement will be provided in Section 7.3. In the following parts of this section, I intended to examine in detail the way *LOOK up*$_3$ and the two potential alternatives, *admire* and *respect*, were used by L1 users, evaluating whether *LOOK up*$_3$ can be replaced by *admire* and *respect*.

7.2.1 Grammatical patterns associated with *LOOK up*$_3$

This section is concerned with the grammatical patterns where *LOOK up*$_3$ frequently appears. Before the examination of grammatical patterns in detail, the first question that

needed to be answered was the number of components this PV contains. In Section 6.2, *LOOK up*$_3$ was regarded as one of the meanings for the verb + particle combination, *LOOK up*. However, it was in fact a three-word expression, *LOOK up to*, whose meaning was associated with the way of showing admiration or esteem towards someone. When this particular meaning made sense in the context, the PV should be regarded as a three-word PV because *LOOK* and *up* failed to present the meaning without *to*. This can be supported by the following example.

> That was a time when the royal family was ***looked up to*** in a manner which could not happen today. (BoE: NB5__020211)

This corpus example revealed the importance of the third component *to*. Since this example used passive voice, the normal structure should be written as:

> That was a time when people ***looked up to*** the royal family in a manner which could not happen today.

The use of passive voice in the example confirmed that *LOOK up to* was a lexicalised unit that functioned as the main verb.

> *That was a time when the royal family was ***looked up*** in a manner which could not happen today.

If *to* was missing, as the example shows, the lexical unit became incomplete and the whole sentence was no longer make sense. As a result, the following examination treated *LOOK up to* as a lexicalised PV and the meaning of this PV was related to the way of showing admiration or esteem towards someone.

Next, I examined the use of *LOOK up to* in the corpus and identified grammatical patterns in which this PV frequently appeared. In the BoE corpus, *LOOK up to* had two different usages. Using the first semantic criterion, the use of *LOOK up to* in the following

examples should be classified as a non-PV combination because the meaning of *LOOK up to* can be predicted by the components.

He ***looked up to*** the sky and waved his cap. (BoE: NB5__020422)

I ***looked up to*** see what had happened. (BoE: BB_cM85__21)

In the two examples, each component took a literal meaning and the meaning of *LOOK up to* was compositional. Consequently, while there were 1678 concordance lines containing *LOOK up to*, only 400 occurrences showing the phrasal use of this combination. It should be noted that the current study took no account of the literal use of *LOOK up to*. Concordance lines containing the literal use were excluded in the examination. The examination of grammatical patterns associated with *LOOK up to* only considered the 400 occurrences in which *LOOK up to* was used as a PV and had a metaphorical meaning.

Among the 400 KWIC lines, **V n** was the most frequent pattern associated with *LOOK up to*, with pronouns such as 'her', 'him', 'you', etc. frequently appearing in the noun slot. Lines 31 to 35 show the use of this simple pattern in the BoE corpus.

31	every actress in the world	**looked up to her**	with a kind of reverence
32	Everyone	**looks up to him.**	He sets the example on the pitch.
33	everybody was	**looking up to him.**	He was the idol of a lot of kids.
34	you could not imagine how I	**looked up to you**	and admired and envied the wit of that
35	especially young people,	**look up to you**	and it's important you are a good role model
36	I suppose she	**looked up to him as**	a kind of father figure.
37	I want my teammates to	**look up to me as**	a leader who can do great things at big
38	In Coull's words, everyone	**looked up to him as**	a respected mentor and leader.
39	And the other police	**look up to her as**	a TV role model.
40	the kids are minorities who	**look up to us as**	role models, and we inspire kids

In addition to **V n**, it seemed that L1 users also used *LOOK up to* in the **V n *as* n** pattern. Lines 36 to 40 provide examples for '*LOOK up to someone as n*'. In **V n *as* n**, pronouns (e.g. 'him', 'me', 'us', etc.) usually appeared in the first noun slot, which was similar

to the **V n** pattern. In terms of the second noun slot, as lines 36 to 40 show, the choice of nouns tended to be individuals who deserved respect due to their influence (e.g. 'a kind of father figure' and 'role model' in lines 36, 39 and 40) or social position (e.g. 'a respected mentor' and 'a leader' in lines 37 and 38). In the BoE corpus, L1 users employed '*LOOK up to someone as n*' to imply the relationship between the people who were being looked up to (i.e. the first noun in **V n *as* n**) and individuals who inspired others (i.e. the second noun). In addition to *LOOK up to*, Francis and Hunston (2000) showed that verbs in the "consider" group, including *consider*, *see*, *regard*, etc., were also able to establish this kind of relationship when they appeared in **V n *as* n**. However, if we use replace *LOOK up to* with verbs in the 'consider' groups, as the following examples show, the replacement cannot be satisfactory. This was because *LOOK up to* had a communicative function of showing admiration and esteem whereas the verb *regard* did not have the same function.

> I want my teammates to ***look up to*** me as a leader who can do great things at big moments. (BoE: NU1__040210)
>
> I want my teammates to ***regard*** me as a leader who can do great things at big moments.

As a result, the examination of *LOOK up to* in the BoE corpus suggested that this PV was frequently used in the **V n *as* n** pattern, suggesting the relationship between the two nouns. From the perspective of pattern grammar, the use of *LOOK up to* was similar to verbs such as *consider, see, regard*, because they were used in the same pattern **V n *as* n**. However, *LOOK up to* cannot be replaced by verbs in the 'consider' group because the PV had a special communicative function of showing admiration and esteem. Verbs in the 'consider' group did not have an equivalent meaning with *LOOK up to*.

7.2.2 **Grammatical patterns associated with *admire* and *respect***

In terms of the two potential alternatives for *LOOK up to*, first of all, corpus evidence showed that the pattern in which *admire* frequently appeared was **V n *for* n**. The following KWIC lines illustrate the use of *admire* in this pattern.

41 I **admire him for** being so confident at his age.
42 people on the boat **admired her for** her courage, and as a good mixer she
43 and I really **admired him for** his competitive spirit.
44 I always **admired Laurance for** his extraordinary business skills
45 and I really **admired them for** their persistence.

When *admire* was used in the **V n *for* n** pattern, the examination of lines 41 to 45 suggested that the two nouns had different functions. The first noun that directly follows the verb often referred to a person (e.g. 'him', 'her', 'Laurance', 'them'), whereas the second noun revealed the characteristics of the person who was worthy of admiration. As the five concordance lines showed, the choice for the second noun could be 'being confident', 'courage', 'competitive spirit', 'extraordinary business skills', 'persistence', etc., which indicated that L1 users preferred to associate *admire* with expressions indicating one's outstanding qualities or talents. As a result, in addition to showing positive regard for others, '*admire someone for n*' was also associated with recognising one's admirable characteristics or talents.

In terms of *respect*, the examination of frequent patterns associated with this individual verb showed that L1 users also used *respect* in the pattern **V n *for* n** frequently. For example,

46 People **respect her for** her abilities and her previous career.
47 supporters saying they **respected him for** acknowledging past mistakes
48 Everyone **respects you for** having done a law degree.
49 And everybody **respected her for** what she did.
50 people would **respect me for** who I am.

As lines 46 to 50 show, in '*respect n for n*', pronouns usually follow the verb directly

and the prepositional phrase introduced by *for* implies the reason why the person is being respected. In other words, L1 users tend to use *respect* and *admire* in a similar manner.

Despite the grammatical similarities between *admire* and *respect*, it seemed that there were some nuances in the meanings of these two verbs. This slight difference can be observed by the choice of nouns in the prepositional phrase (i.e. the second noun slot in **V n *for* n)**. Unlike expressions that were associated with one's outstanding qualities or talents in '*admire n for n*', L1 users tended to choose expressions indicating one's efforts, achievements, or authority in '*respect n for n*'. For example,

Everyone ***respects*** you for having done a law degree (BoE: NB1__040518)

Most people ***respect*** Steve for what he has achieved. (BoE: NA1__040822)

In the first example, the reason why the person deserves everyone's respect is due to the efforts of 'having done a law degree'. Also, in the second example, the *wh*-clause, 'what he has achieved', implies the superior achievements of the person. Therefore, because of the specific choice in the second noun slot, it can be argued that the meaning of '*respect n for n*' was associated with acknowledging one's effort and achievements.

This section has examined grammatical patterns associated with *admire* and *respect* in the BoE corpus. The findings showed that the two verbs were used in a similar way, and both of them frequently appeared in the pattern **V n *for* n**. When *admire* and *respect* were used in this pattern, pronouns usually appeared in the first noun slot and the second noun tended to be associated with one's outstanding qualities or achievements. In terms of the semantic use, although both verbs implied positive regard for someone, corpus evidence suggested that '*admire n for n*' and '*respect n for n*' had a slight difference.

The semantic nuance can be revealed by the lexical choices in the second noun. L1 users tended to use '*admire n for n*' to show their admiration for one's outstanding qualities or talents, whereas they tended to use '*respect n for n*' to express the acknowledgement of one's achievements or efforts.

7.2.3 The failure of replacing *LOOK up to* with *admire* and *respect*

In the previous two sections, I have discussed the frequent patterns associated with the target PV, *LOOK up to*, and two potential alternatives, *admire* and *respect*. The findings from the BoE corpus showed that the PV was frequently used in **V n** and **V n *as* n** patterns, whereas *admire* and *respect* frequently appeared in **V n** and **V n *for* n** patterns. Based on the findings, it seemed that *LOOK up to* could be replaced by *admire* or *respect* when they were used in the same grammatical pattern (i.e. **V n**). This claim, however, lacked evidence to show that *LOOK up to* was semantically interchangeable with the two verbs when they appeared in **V n** pattern. As a result, this section firstly compared the semantic features between *LOOK up to* and the two verbs. In addition, this section also discussed the grammatical differences between them, examining whether L1 users used *LOOK up to* in a way that was similar to *admire* and *respect* (i.e. whether *LOOK up to* can be used in the **V n *for* n** pattern).

7.2.3.1 Semantic differences

Corpus data showed *LOOK up to*, *admire*, and *respect* can appear in the simple pattern **V n.** When they were used in this pattern, they were frequently followed by pronouns, such as *her, him, you*, etc. Lines 51 to 55 below demonstrate the use of *LOOK up to* in the **V n** pattern and suggest the semantic meaning of this PV.

51 Milton, as a small boy, had **looked up to his father** as small boys usually do.

52 You have guys that **look up to you** as a bit of a leader and you've got to show them something and give them something.

53 Ion was much older than me and I used to **look up to him**, but once he realised I was a much better player than him he started to resent me and ignore me.

54 He was loved by everyone, especially the young kids. They all **looked up to him** because he looked out for them.

55 Did you **look up to the older ones** or fear them? Were they more or less successful than you?

A close examination of lines 51 to 55 showed that when L1 users employed '*LOOK up to someone*', they tended to suggest a hierarchical relationship in the context. The person being looked up to was considered superior based on factors such as age, power, or social status. For example, 'Milton *looked up to* his father' in line 50 suggested the relationship between sons and fathers, implying that fathers were more powerful than sons in a family. Similarly, the contextual information in line 54 also indicated this kind of relationship. In line 54, 'the young kids *looked up to* him' suggested that 'he' was older, more experienced, and acted as a protector who 'looked out for the young kids'.

Based on the analysis of *LOOK up to* in the **V n** pattern, Table 7.2 below summarises the semantic meaning of *LOOK up to* and integrates the findings of *admire* and *respect* from the BoE corpus (see Section 7.2.2). As Table 7.2 shows, *admire* usually described positive regards towards one's outstanding qualities or talents, and *respect* was used to express the acknowledgement of one's achievements or efforts. When the two verbs appeared in the **V n** pattern, L1 users tended to associate superiority with factors such as one's extraordinary skills or great achievements. *LOOK up to*, however, established a hierarchical relationship in the same pattern, connecting one's superiority with factors such as age, power, social status, etc. As a result, due to the semantic nuances between *LOOK up to* and the two one-word verbs when they all appeared in the **V n** pattern, it

can be argued that *LOOK up to* cannot be replaced with neither of the two verbs adequately.

Table 7.2 Semantic meanings and examples of *LOOK up to*, *admire*, and *respect*.

	Semantic meanings	**Corpus examples**
LOOK up to	To establish a hierarchical relationship based on factors such as age, power, social status, etc.	Milton, as a small boy, had ***looked up to*** his father as small boys usually do. (BoE: BA_Qf022206)
admire	To show positive regards towards one's outstanding qualities or talents	I always ***admired*** Laurance for his extraordinary business skills. (BoE: NC4__040712)
respect	To acknowledge one's achievements or efforts	Everyone ***respects*** you for having done a law degree. (BoE: NB1__040518)

7.2.3.2 Grammatical differences

The previous section discussed the failure of replacing *LOOK up to* with *admire* and *respect* because of the semantic nuances between the PV and the one-word verbs. In addition to the semantic differences, the grammatical patterns in which the PV and the two verbs frequently appeared were also different. In the BoE corpus, L1 users preferred to use '*admire n for n*' or '*respect n for n*' rather than '*LOOK up to n for n*'. It was rare to find concordance lines showing the use of *LOOK up to* in **V n *for* n**.

```
56    younger lads, who obviously look up to him for what he has achieved in the game
57                    people have looked up to us for years.
58                         I have looked up to you for quite some time and am a great fan

44                       I always admired Laurance for his extraordinary business skills
48                       Everyone respects you for having done a law degree.
```

Lines 56, 57 and 58 demonstrated two types of usage when *LOOK up to* was used in **V n *for* n**. Compared with examples that illustrate '*admire n for n*' (line 44) and '*respect n for n*' (line 48), it can be observed that lines 57 and 58 were completely different. As lines 44 and 48 showed, the prepositional phrases, 'for his extraordinary business skills' and 'for having done a law degree' implied the reasons why someone was being

admired or respected by others. However, in lines 57 and 58, '*looked up to* us for years' and '*looked up to* you for quite some time' suggested the duration over which the speakers admired or respected someone, emphasizing that this admiration or respect lasted for a long time. The examination provided evidence that *LOOK up to* was rarely used in the pattern that *admire* and *respect* frequently appeared.

In terms of **V n *as* n**, the pattern where *LOOK up to* frequently appeared, corpus evidence showed that it was possible for *admire* and *respect* to appear in **V n *as* n**. Therefore, I also examined whether *LOOK up to* and the two verbs had a similar use when they all appeared in the **V n *as* n** pattern. When *LOOK up to* appeared in **V n *as* n**, the findings showed that this PV was followed by pronouns directly. Expressions indicating one's influence or social position (e.g. 'role model', 'a leader', 'a kind of father figure') often appeared in the second noun slot (see Section 7.2.1). In order to examine whether *admire* and *respect* could substitute the PV in this pattern, I retrieved the use of '*admire n as n*' and '*respect n as n*' in the BoE corpus. The following concordance lines demonstrate the use of *admire* and *respect* in **V n *as* n**.

59 Peggy **admired her as** someone she herself could never be.
60 I **admire him as** an actor because he is completely generous in his work.
61 They loved him as a friend and **admired him as** a songwriter.
62 I told him I **admire him as** a teammate and appreciate how he has handled this.
63 I do **admire him as** a poet but we haven't really associated with each other all that much.

64 I **respect him as** a fighter and as a champion and think he would be a quality opponent
65 I want him to **respect me as** a career woman.
66 I want to get to where others in the field **respect me as** a good gardener.
67 I find the best way to relate to teenagers is to treat them and **respect them as** adults.
68 It helps her to understand you and it will encourage her to **respect you as** an individual.

Lines 59 to 63 illustrate how L1 users employed *admire* in **V n *as* n**, and lines 64 to 68 showed the use of *respect* in this pattern. As the ten concordance lines showed, pronouns also followed the verb directly, but L1 users tended to choose expressions

such as ‘an actor’, ‘a songwriter’, ‘a fighter’, ‘a good gardener’, etc. in the second noun slot. It seemed that these expressions were concerned with specific identities or professions. In other words, the meaning of ‘*admire n as n*’ and ‘*respect n as n*’ tended to be associated with recognising or appreciating one’s achievements in a specific identity or profession.

> I want my teammates to ***look up to*** me as a leader who can do great things at big moments. (BoE: NU1__040210)
>
> I ***admire*** him as an actor because he is completely generous in his work. (BoE: BU_Ym941237)
>
> I ***respect*** him as a fighter and as a champion and think he could be a quality opponent (BoE: NBA__040731)

The three examples were chosen to show the differences between ‘*look up to n as n*’ and ‘*admire/respect n as n*’. In the first example, ‘*look up to* me as a leader’ indicated the superior social position of the respected person. However, there was no indication of such superiority in the second and third examples. L1 users intended to use ‘*admire* him as an actor’ and ‘*respect* him as a fighter and as a champion’ to convey the approval of one’s professional achievements as ‘an actor’ or ‘a fighter’. As a result, the closer examination suggested that, while it was possible for *admire* and *respect* to replace *LOOK up to* in the pattern **V n *as* n**, the replacement cannot be satisfactory because the two individual verbs conveyed a meaning that was different from the PV.

Furthermore, there was an additional point showing the grammatical differences between *LOOK up to* and the two verbs. In the BoE corpus, *admire* and *respect* can be used in passive voice, but it was less likely to find the passive voice of *LOOK up to*. Table 7.3 shows the frequency of both active and passive voice for *LOOK up to*, *admire,*

and *respect* in the BoE corpus.

Table 7.3 The active and passive voice for *LOOK up to*, *admire,* and *respect*

	Constructions	Frequency	Freq. per million words	Percentage (%)
LOOK up to	*LOOK up to someone*	400	0.63	90.09
	be looked up to	44	0.06	9.91
admire	*admire someone for n*	259	0.36	68.16
	someone is admired for n	121	0.17	31.84
respect	*respect someone for n*	199	0.28	68.62
	someone is respected for n	91	0.13	31.38

As Table 7.3 shows, there were 121 concordance lines showing the passive voice form '*someone is admired for n*', accounting for approximately one third of the total occurrences. Similarly, the passive voice form '*someone is respected for n*' also accounted for one third of the occurrences. However, there were only 44 concordance lines showing the passive voice of *LOOK up to*, which was substantially less frequent than the active voice of this PV. One possible reason for this grammatical difference could be the nature of English PVs. In English, some PVs do not have the passive voice because transforming from the active voice to the passive voice may reduce clarity, leading to awkward or cumbersome use of language (Biber et al., 1999).

To sum up, based on the examination of *LOOK up to* in the BoE corpus, it can be argued that this PV did not have a satisfactory one-word replacement. The findings of *admire* and *respect* showed that *LOOK up to* and the two potential one-word alternatives had semantic and syntactic differences. L1 users seldom employed *LOOK up to* in the **V n *for* n** pattern, but it is a common pattern associated with *admire* and *respect*. In terms of the semantic differences, when this PV and the two verbs appeared in the same grammatical pattern (e.g. **V n**), the meaning of '*LOOK up to someone*' indicated a

hierarchical relationship and suggested one's superiority based on factors such as age, power, and social status. In terms of '*admire someone*' and '*respect someone*', there was no evidence to show that they indicated this kind of relationship in the context. Moreover, while corpus evidence showed that both *LOOK up to* and the two verbs can be used in **V n *as* n**, the substitution remained unsatisfactory. L1 users had different semantic choices in the second noun slot. The semantic meaning of '*admire/respect n as n*' was associated with recognising or appreciating one's achievements in a specific profession, whereas '*LOOK up to someone as n*' tended to reflect one's moral or social superiority.

7.3 The situations of replacing PVs with a multi-word combination

In the previous two sections, the examinations of *LOOK for*$_1$ and *LOOK up to* provided evidence to support the argument that using one-word verb to replace PVs can be challenging. Grammatical differences played an important role in the substitution. The ignorance of grammatical patterns associated with PVs and their potential one-word alternatives can result in an unsuccessful replacement. When it was difficult to find one-word alternatives for PVs, Cornell (1985) implied that the potential alternative could be other PVs or multi-word combinations that were consistent in meaning. For example,

69 The man was due to arrive in the next few days to **take over** the burden of the remaining children's lessons.

70 He knew how to **get on with** people and what makes people do what they do.

In the two examples, it was hard to find one-word equivalents to replace *take over* and *get along with*. Possible rephrasing for lines 69 and 70 could be:

(69) The man was due to arrive in the next few days and he would be responsible for the remaining children's lessons.

(70) He knew how to have a good relationship with people and what makes people do what they do.

Multi-word expressions, 'be responsible for' and 'have a good relationship with', were employed to replace *take over* and *get on with* without changing the meaning of the original sentences. This rephrasing practice suggested that PVs without one-word equivalents were possible to be substituted by multi-word expressions. As a result, this section focuses on situations when PVs only have multi-word alternatives. This section selected two PVs, *LOOK after* and *LOOK out for*. Both of them seemed to be synonymous with a multi-word expression, *TAKE care of*. By examining the use of these three phrases in the corpus, this section discusses whether *TAKE care of* can substitute *LOOK after* and *LOOK out for* successfully.

7.3.1 The replacement between *LOOK after* and *TAKE care of*

First of all, *LOOK after* was classified as a PV in the current study because the meaning of this combination cannot be predicted by the components (see lines 71 to 75).

71	in the last 18 months to	**look after**	her two children, aged five and six.
72	If you're	**looking after**	someone with a disability, you may want to leave
73	rather desperate: to	**look after**	the poor and the unemployed and to help them
74	not go to doctors or	**look after**	their own health, but also lack access to doctors
75	The bank agrees to	**look after**	your money and provide you with a good customer

As the five concordance lines show, in the BoE corpus, the most frequent grammatical pattern associated with this PV was the **V n** pattern. Nouns following *LOOK after* referred to people with disability or vulnerability (e.g. 'children', 'someone with a disability', etc.), or objects that are highly valuable (e.g. 'their own health', 'your money', etc.). In terms of *TAKE care of*, lines 76 to 80 were extracted from the BoE corpus to show the use of this multi-word combination.

76 he can trust his team to **take care of** business in his absence
77 was perfectly capable of **taking care of** herself, either in harmony or in opposition to
78 the problem will be **taken care of** in the next few years.
79 jobs in our schools, **taking care of** our kids, making sure they're safe
80 I think if we can **take care of** their basic needs and then get it all coordinated

As lines 76 to 80 show, L1 users used *TAKE care of* in the **V n** pattern, which was similar to *LOOK after*. The word choice in the noun slot also referred to people with vulnerability, such as 'kids' (line 79). It seemed so far that there was no significant difference in the syntactical use of *LOOK after* and *TAKE care of*. However, it might be too early to draw the conclusion that *LOOK after* can be replaced with *TAKE care of*. A closer examination of concordance lines including *LOOK after* and *TAKE care of* suggested that there were subtle differences between the two combinations.

Evidence from the corpus suggested that *TAKE care of* tended to co-occur with expressions which had no association with value, such as 'the problem' (line 78) and 'their basic needs' (line 80). When *TAKE care of* was followed by these expressions, the meaning seemed to be associated with providing essential support or assistance to solve the problem or fulfill the needs. However, corpus evidence of *LOOK after* showed that this PV was used in a different way from *TAKE care of*. The choice of nouns following *LOOK after* was restricted. In the BoE corpus, it was rare to find this PV to co-occur with expressions such as 'the problem'. As a result, the findings from the corpus showed that *TAKE care of* had a wider choice of nouns, whereas the nouns that followed *LOOK after* tended to be restricted to those associated with vulnerability and value (e.g. '*look after* her two children' in line 71 and '*look after* their own health' in line 74). This suggested that it was possible to use *TAKE care of* to replace *LOOK after*, but this replacement was not bilateral.

The bank agrees to ***look after*** your money and provide you with a good

customer service. (BoE: MBX_____78)

The bank agrees to ***take care of*** your money and provide you with a good customer service.

The problem will be ***taken care of*** in the next few years. (BoE: NU5__040303)

*The problem will be ***looked after*** in the next few years.

7.3.2 The replacement between *LOOK out for* and *TAKE care of*

Moving on to *LOOK out for*, this section firstly discusses why this three-word combination was regarded as a PV in the current study, and then examined whether *LOOK out for* can be replaced with *TAKE care of*.

7.3.2.1 The phrasal identity of *LOOK out for*

The following five examples provide evidence to support that *LOOK out for* should be regarded as a PV with a non-compositional meaning.

81	the 12-year-old was always	**looking out for** other kids in the neighbourhood.
82	job is really to	**look out for** consumers by really taking a hard line
83	neighbours' in those days	**looked out for** each other more than they tend to do
84	begged her brother-in-law to	**look out for** her son. She'd not seen him since
85	I know those people will be	**looking out for** them when they're in those boat.

The examination of the KWIC lines suggested that the meaning of *LOOK out for* seemed to be associated with taking care of someone or something, such as '*looking out for* other kids' (line 81) and '*look out for* her son' (line 84). There was no clue showing the association between this meaning of *LOOK out for* and the three components. Since *LOOK out for* had an non-compositional meaning in the given examples, this combination should be labelled as a PV using the first semantic criterion (see Section 5.1.1).

It may be claimed that *LOOK out for* is a combination in which a two-word verb, *LOOK out* is combined with a preposition, *for*. This claim regards *LOOK out* as a lexical unit which is followed by *for*. The following concordance lines show the use of *LOOK out* in the BoE corpus.

86	"Kate!	**Look out**!"	Greg's voice bellowed out into the wind.
87	"Hey!	**Look out**!"	shouted Patty, as the front end lifted off the ground
88	I	**looked out**	and saw this sea of raised hands, right back to
89	Selby	**looked out**	the front door and saw something strange in the field
90	When I	**looked out**	the window I could see the boats in the harbour

The use of *LOOK out* in lines 86 and 87 was different from the use in the other three lines. In lines 86 and 87, the meaning of *LOOK out* was associated with giving a warning, which cannot be predicted by *LOOK* and *out*. In other words, *LOOK out* had a non-compositional meaning in lines 86 and 87 and should be classified as a PV. However, in lines 90, 91 and 92, *LOOK out* cannot be regarded as a PV because the literal meaning of each component helped to interpret the meaning of this combination. Moreover, in order to explore the role that *for* played when it joined *LOOK out*, I investigated the occurrences when *LOOK out* was followed by other prepositions that frequently co-occurred with *LOOK out*, including *at*, *on* and *over*. Lines 91 to 96 illustrate the use of these combinations in the BoE corpus.

91	While Emily	**looked out**	at the audience, she could hear the voice of the woman
92	He	**looked out**	at the mountains in the distance.
93	one can	**look out**	on one side of the ship and see an eternity of still
94	he could	**look out**	on the river bank and be certain he had played
95	She	**looked out**	over the sea of people and smiled.
96	I	**looked out**	over a tiny garden surrounded by high brick walls

When *LOOK out* was followed by *at*, *on* and *over*, it was less evident that these combinations had the potential to be a PV. For example, the meaning of *LOOK out at* in lines 91 and 92 were compositional, associated with the literal meaning of each component. Similarly, the examination of '*look out* on the river bank' in line 94 and

'*looked out* over a tiny garden' in line 96 also showed that the meaning of *LOOK out* was compositional and cannot be classified as a PV. Therefore, when *LOOK out* was combined with other prepositions such as *at*, *on*, and *over*, it was not possible to claim that these combinations were used as a PV. The use of *LOOK out* in *LOOK out for* tended to be distinctive because it had a non-compositional meaning. Also, *for* played an essential role in *LOOK out for* as *LOOK out* was less likely to be a lexicalised unit without *for*.

7.3.2.2 Grammatical patterns associated with *LOOK out for*

After confirming the phrasal identity of *LOOK out for*, I investigated the use of this three-word PV in the BoE corpus and considered potential replacements for it. In the previous section, lines 81 to 85 showed that L1 users frequently used *LOOK out for* in the **V n** pattern. The following concordance lines provide additional evidence to show the use of this pattern.

```
81     the 12-year-old was always looking out for other kids in the neighbourhood.
82               job is really to look out for consumers by really taking a hard line
83      neighbours' in those days looked out for each other more than they tend to do
84   begged her brother-in-law to look out for her son. She'd not seen him since
85    I know those people will be looking out for them when they're in those boat.

97        to make friends and can look out for familiar faces on his first day
98            Sara must have been looking out for him, because she opened the door before he
99                           I'll look out for you in the Three Fiddlers at the weekend
100                             I looked out for you at Paddington, but I didn't see you
```

In the additional lines (i.e. lines 97 to 100), *LOOK out for* was also followed by nouns. Since *LOOK out for* and *TAKE care of* can be used in the same pattern (i.e. **V n**), it might be successful to replace *LOOK out for* with *TAKE care of*. For instance,

> I know those people will be ***looking out for*** them when they're in those boats. (BoE: NU4__030512)
>
> I know those people will be ***taking care of*** them when they're in those boats.

In the given example, *LOOK out for* can be substituted by *TAKE care of*, and the replacement did not change the structure and meaning of the original sentence. However, a closer examination suggested that *TAKE care of* failed to substitute *LOOK out for* in lines 97 to 100 due to the polysemous feature of *LOOK out for.* In the BoE corpus, *LOOK out for* had multiple meanings, one of which was illustrated by lines 81 to 85. Another usage was suggested by lines 97 to 100.

I ***looked out for*** you at Paddington, but I didn't see you. (BoE: BB_cF86__25)

* I ***took care of*** you at Paddington, but I didn't see you.

As this example showed, the meaning of '*look out for* you' seemed to be associated with 'meeting you' due to the contextual information 'at Paddington' and 'I didn't see you'. In this instance, *LOOK out for* did not imply the idea of 'taking care of someone', and therefore it cannot be replaced with *TAKE care of*. In addition, evidence from the BoE corpus revealed that this PV had a third usage that was different from the other two. For example,

101	When buying plastics	**look out for** the following recyclable types
102	but we are also	**looking out for** public safety and any liabilities for
103	It is always worth	**looking out for** special fixed-menu offers as these can be
104	families as a whole and	**looking out for** danger signs of relationship problems
105	asking Londoners to	**look out for** unattended or suspicious bags and to dial 999

In lines 101 to 105, the interpretation of '*LOOK out for something*' seemed to be associated with 'paying attention to something that looks special or unusual', such as 'danger signs' (line 104) and 'unattended or suspicious bags' (line 105).

We are also ***looking out for*** public safety and any liabilities for the American taxpayer. (BoE: NU5__031011)

We are also ***taking care of*** public safety and any liabilities for the American taxpayer.

In the given example, while it was grammatically correct to replace '*looking out for*

public safety' with '*taking care of* public safety', there were subtle semantic differences between them. When L1 users employed '*looking out for* public safety', they tended to convey the idea that potential threats or risks were concerned to ensure public safety. However, it seemed that '*taking care of* publish safety' implied the responsibility for ensuring the safety, without the indication of paying attention to potential threats or risks. This semantic nuance led to an unsuccessful replacement between *LOOK out for* and *TAKE care of*.

To sum up, this section paid attention to situations in which PVs did not have one-word equivalents. *LOOK after* and *LOOK out for* were two PVs that cannot be substituted by one-word verbs but had potential multi-word equivalents. The findings from the BoE corpus showed that L1 users usually employed the two PVs in the **V n** pattern. When they appeared in the **V n** pattern, *TAKE care of* could be a potential replacement for these two PVs without the consideration of semantic nuances. However, the examination of the semantic use suggested that *TAKE care of* could not be a satisfactory replacement for *LOOK after* and *LOOK out for*. Corpus evidence showed that *LOOK after* tended to be followed by people with vulnerability (see lines 72 and 73) or objects that are highly valuable (see lines 74 and 75), whereas there was no explicit evidence showing that *TAKE care of* had a similar semantic use.

In terms of *LOOK out for*, this PV has multiple meanings in the BoE corpus. Because of the semantic complexity of this PV, it was hard to find a perfect alternative for *LOOK out for*. Table 7.4 below provides information about the multiple meanings of *LOOK out for* and suggests the extent to which it can be replaced with *TAKE care of*.

Table 7.4 Semantic complexity of *LOOK out for* and the replacement with *TAKE care of*

	Pattern	Examples	Replacement with *TAKE care of*
LOOK out for	**V n**	I know those people will be ***looking out for*** them when they're in those boats. (BoE: NU4__030512)	Yes
		I ***looked out for*** you at Paddington, but I didn't see you. (BoE: BB_cF86__25)	No
		We are also ***looking out for*** public safety and any liabilities … (BoE: NU5__031011)	No

As Table 7.4 shows, corpus evidence suggested that there was one situation in which *LOOK out for* can be replaced with *TAKE care of*. However, *LOOK out for* was used in a different way from *TAKE care of* in '*looked out for* you at Paddington' and '*looking out for* public safety'. Because of the semantic nuances shown by the examples, *LOOK out for* cannot be replaced with *TAKE care of* successfully.

7.4 Summary

This chapter examined the replacement of PVs from the perspective of pattern grammar and discussed the extent to which an English PV can be replaced successfully in three different situations. The results suggested that finding an appropriate replacement for English PVs tended to be challenging. Section 7.1 was concerned with the situation when a PV has a one-word equivalent (e.g. *LOOK for*$_1$ vs *search*). The analysis showed that the patterns where *search* frequently appeared (i.e. **V n *for* n** and **V *for* n**) were different from the patterns where *LOOK for*$_1$ appeared (i.e. **V n *in* n**). The different grammatical patterns resulted in the failure of the replacement between *LOOK for*$_1$ and *search*. Section 7.2 focused on a polysemous PV, *LOOK up*, and showed that it was impossible to find a one-word verb to replace this polysemous PV. Each meaning of

LOOK up had a potential replacement. However, whether the replacement could be satisfactory still relies on the consideration of both lexical and grammatical patterns. Finally, Section 7.3 focused on the situation in which PVs did not have one-word equivalents. This section showed the unsuccessful replacement between PVs and multi-word alternatives. From the perspective of grammatical patterns, it seemed that *LOOK after* and *LOOK out for* can be substituted by *TAKE care of* because they all appeared in the **V n** pattern. However, if we looked into the semantic nuances, the results showed that *TAKE care of* failed to be a satisfactory replacement for both *LOOK after* and *LOOK out for*. As a consequence, when L2 learners replace PVs with one- or multi-word verbs, they need to have a clear understanding of two aspects: (1) the grammatical patterns associated with PVs and potential alternatives, and (2) the semantic nuances between them.

Based on the examination of replacements in this chapter, the findings could draw L2 learners' attention to the use of thesaurus. Many L2 learners often avoid using PVs and prefer finding a near synonym in thesaurus. However, relying on thesaurus may cause problems. One of the negative effects of thesaurus is the lack of information for the syntactic use. For example, *LOOK for* and *search* are a pair of near synonyms, but they are used in different grammatical patterns. Without the awareness of this grammatical difference, learners are likely to make mistakes. Another negative effect of relying on thesaurus is associated with the semantic nuances between near synonyms. *TAKE care of* fails to be a satisfactory replacement for *LOOK out for* because L1 users employ *LOOK out for* to convey specific meanings that seem to be implicit when this PV is considered in isolation. In thesaurus, learners may find a list of synonyms. However, the lexical and grammatical patterns associated with PVs, such as collocation, are unable to know.

Chapter 8 The use of PVs in Chinese learners' written texts

In the previous two chapters, I have examined the use of specific PVs in the BoE corpus, with a particular focus on the lexico-grammatical patterns associated with these PVs. The results from the BoE corpus provided insights into L1 users' use of English PVs and showed the nuances between PVs and synonymous one-word verbs. After understanding the lexico-grammatical patterns associated with specific PVs in the BoE corpus, this chapter moves on to the second and the third research questions of the study, investigating how Chinese EFL learners use PVs in their writing (see Section 4.1). The aim of this chapter is to compare the use of frequent PVs in the Chinese learner corpus and the use in the reference corpus. As mentioned in Section 4.2.1, the current study did not use the BoE corpus for comparison. As a result, in this chapter, corpus evidence showing how L1 users used frequent PVs in their writing were extracted from the reference corpus.

This chapter is divided into three sections. Firstly, Section 8.1 is an overview of Chinese EFL learners' use of English PVs in written texts. In this section, I will focus on frequent PVs with two or three components, comparing the frequency of PVs in the Chinese learner corpus and the reference corpus. Section 8.2 provides insights into the use of specific PVs in the learner corpus. It aims to show Chinese learners' performance when they employ PVs such as *LOOK up* and *LOOK forward to*. Finally, Section 8.3 is concerned with learners' choice between PVs and synonymous one-word verbs. In this section, I will compare the use of *LOOK for* and *search* in the learner corpus and examine learners' performance when they use this pair of near synonyms in their writings.

8.1 The use of frequent PVs in the Chinese learner corpus

8.1.1 Frequency and distribution of frequent PVs in the Chinese learner corpus

Frequency data and statistical association data are found to be reliable predictors of learners' knowledge of vocabulary (Gries and Durrant, 2021). Therefore, the current study regarded the overall frequency of frequent PVs in the Chinese learner corpus as an important indicator of learners' performance of using PVs. As mentioned in Section 1.4, the study focused particularly on the 100 frequent PVs identified by Gardner and Davis (2007). The authors observed that these PVs had a relatively high coverage in the BNC, accounting for more than half (50.4%) of the total occurrences of PVs in the corpus (Gardner and Davis, 2007). Because of the high coverage, this list of frequent PVs is considered a useful resource for EFL learners (Garnier and Schmitt, 2015). In the list, Gardner and Davis's (2007) found the majority of PVs (86 out of 100 PVs) were constructed by 20 verbs and 8 particles. Therefore, in this section, the 100 PVs were divided into 9 groups: 86 of them were categorised by the 8 particles and the remaining 14 PVs were included in the 'others' group (see Table 8.1).

Table 8.1 The frequency of 100 frequent PVs in the Chinese learner corpus

Particle	Examples	Frequency	Freq. per million
out (19)	*FIND out, PICK out, etc.*	1307	453.44
up (15)	*BRING up, COME up, etc.*	1921	666.45
back (12)	*HOLD back, PUT back, etc.*	329	114.14
down (10)	*BREAK down, SET down, etc.*	182	63.14
in (8)	*GIVE in, TAKE in, etc.*	234	81.18
off (8)	*GET off, SET off, etc.*	187	64.88
on (8)	*CARRY on, TURN on, etc.*	539	187.00
over (4)	*TAKE over, TURN over, etc.*	43	14.92
others (14)	*BRING about, GO through, etc.*	330	114.49
Total		**5072**	**1759.63**

(*Notes*: a. out (19) means there are 19 PVs in the 'out' group in the Gardner and Davis's list; b. the complete list of the 100 frequent PVs is given in Appendix 1.)

Table 8.1 provides both the observed frequency and the normalised frequency of frequent PVs in each group. The data in Table 8.1 indicated that PVs in the 'out' and 'up' groups were frequent in their writing, whereas PVs in the 'over' group were less frequent. The corpus evidence may be associated with learners' familiarity with these items. The low frequency of PVs in the 'over' group might indicate that learners were not familiar with them. With corpus data, it should be noted that we cannot directly build connection between textual evidence and learners' knowledge of PVs. In other words, the conclusion needs to be corroborated with further non-textual evidence such as learners' performance data. The low frequency of PVs in the 'over' group was probably due to the number of PVs included in this group. The number of PVs in the 'out' and 'up' groups (19 PVs and 15 PVs respectively) was more than that in the rest seven groups. However, if we focused on the occurrences of PVs in the 'out' group and the 'up' group, the observation showed that PVs in the 'up' group occurred more frequently (1921 occurrences, 666.45 per million words) than that in the 'out' group (1307 occurrences, 453.44 per million words). To put it another way, more PVs included in the 'out' group did not indicate a higher frequency of this group in the learner corpus. Chinese EFL learners seemed to be more familiar with frequent PVs constructed by *up* rather than *out*. Similar trend can be found in the 'down' group and the 'on' group. While there were more PVs included in the 'down' group (10 PVs), the normalised frequency of this group (63.14 per million words) was three times lower than the normalised frequency of the 'on' group (187.00 per million words).

In terms of the overall frequency of the 100 frequent PVs, Table 8.1 shows that they occurred 5072 times (1759.63 per million words) in the Chinese learner corpus. To be more specific, the normalised frequency suggested that Chinese EFL learners employed

less than two PVs out of 1000 words in their writing. The distribution of frequent PVs in the learner corpus was lower than the estimation made by Gardner and Davis (2007). Gardner and Davis (2007) estimated that learners may encounter on average two PVs per 300 English words they were exposed to. However, it should be noted that the estimated distribution proposed by Gardner and Davis (2007) was based on both written and spoken texts in the BNC. The comparison seemed invalid because spoken texts were absent from this study. As a result, in order to examine whether the 100 frequent PVs had a lower distribution in the learner corpus (i.e. whether Chinese EFL learners underused the frequent PVs in their writing), the study also compared the frequency and distribution of these PVs in the reference corpus and incorporated with the chi-square test to assess the significance of the results. Table 8.2 provides the results for the comparison between the two corpora and the frequency of PVs in each subset group.

Table 8.2 The frequency of 100 frequent PVs in the reference corpus and the Chinese learner corpus

	Frequency		Freq. per million	
	Reference corpus	Learner corpus	Reference corpus	Learner corpus
Total	**9579**	**5072**	**2241.43**	**1759.63**
out (19)	3210	1307	751.12	453.44
up (15)	2028	1921	474.54	666.45
back (12)	788	329	184.39	114.14
down (10)	604	182	141.33	63.14
in (8)	585	234	136.89	81.18
off (8)	321	187	75.11	64.88
on (8)	1247	539	291.79	187.00
over (4)	291	43	68.09	14.92
others (14)	505	330	118.17	114.49

Overall, the results showed that the 100 frequent PVs had a lower distribution in Chinese learners' writing compared to that in L1 users' writing. This underuse of frequent PVs in the learner corpus was statistically significant ($\chi^2 = 168.74$, $p < 0.001$).

As Table 8.2 shows, in the reference corpus, the normalised frequency for PVs was 2241.43 times per million words, which was higher than the normalised frequency observed in the learner corpus (1759.63 per million words). The normalised frequency shown in Table 8.2 also indicated the lower distribution of PVs in learners' writing.

In addition, the distribution of PVs in each group showed that learners seemed to have difficulty using certain PVs in their writing, especially PVs in the 'down', 'in' and 'over' groups. As Table 8.2 shows, in the Chinese learner corpus, the normalised frequency of the 'down' group was 63.14 times per million words, which was two times lower than the normalised frequency of this group found in the reference corpus (141.33 per million words). Such dramatic difference between the two corpora was also true for the 'over' group. In the learner corpus, the distribution of PVs in the 'over' group (14.92 per million words) was substantially lower than that in the reference corpus (68.09 per million words). The underuse of PVs in the 'down', 'in' and 'over' groups may suggest a lack of familiarity or a preference for alternative constructions. While this could suggest that Chinese EFL learners are less familiar with PVs in these groups and may find them challenging to use in writing, further evidence, such as error analysis or quiz results, would be needed to confirm this interpretation.

In the reference corpus, the most frequent group of PVs was the 'out' group (751.12 occurrences per million words), which was followed by the 'up' group, occurring 474.54 times per million words. However, this was different from the trend observed in the learner corpus. In the learner corpus, the normalised frequency of the 'up' group was 666.45, which was even higher than that in the reference corpus. While the high frequency of the 'up' group in the learner corpus could reflect learners' familiarity with

specific PVs, this disproportionately higher distribution may indicate the overuse of PVs in the 'up' group in Chinese EFL learners' writing (see the examination of *LOOK up* in Section 8.2).

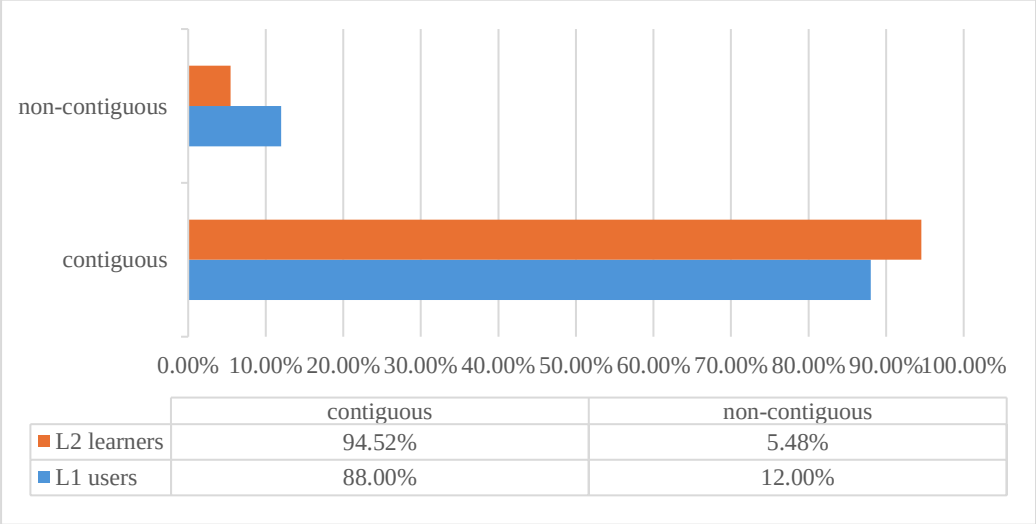
Based on the statistical results of the 100 frequent PVs in the two corpora, it can be argued that the use of these PVs in the learner corpus differed significantly from that in the reference corpus. Chinese EFL learners tended to underuse PVs in their writing, especially PVs from the 'down', 'in' and 'over' groups. The lower distribution of the three groups suggested that Chinese EFL learners tended to be less familiar with PVs such as *TAKE down, GIVE in, TURN over,* etc. Therefore, they need more exposure to these PVs in order to improve their familiarity and proficiency. While two-word PVs were generally less frequent in the learner corpus, it was noteworthy that PVs in the 'up' group had a higher distribution in learners' writing. Although this high distribution suggested that Chinese EFL learners were familiar with PVs in the 'up' group, it seemed to be disproportionate, as the distribution was even higher than that in the reference corpus. I will revisit this point in Section 8.2, where qualitative analysis of the use of PVs in the 'up' group will be presented.

8.1.2 Contiguous and non-contiguous forms of frequent PVs in the Chinese learner corpus

The movement of the particle is a typical feature for English PVs (see Section 2.1.3). In the Gardner and Davis's (2007) list of top 100 most frequent PVs, more than half of them are separable PVs. More precisely, 55 out of 100 frequent PVs allow the object to separate the verb and the particle (e.g. *bring up the kid = bring the kid up*). Therefore,

this section presents the findings on the use of both contiguous and non-contiguous forms in the two corpora, with a particular focus on the use of the non-contiguous form in learners' writing.

Figure 8.1 shows the proportions of the two forms in the reference corpus and the Chinese learner corpus. As Figure 8.1 shows, the percentage of the contiguous form in both corpora was substantially larger than the percentage of the non-contiguous form. This suggested that both L1 users and Chinese EFL learners were inclined to use the contiguous form of PVs in their writing.



	contiguous	non-contiguous
L2 learners	94.52%	5.48%
L1 users	88.00%	12.00%

Figure 8.1 The contiguous and non-contiguous forms of separable PVs in the Chinese learner corpus and the reference corpus.

If we focused on the proportion of the non-contiguous form, the overall percentage observed in Figure 8.1 indicated that Chinese EFL learners were able to use the non-contiguous form in their writing. However, the non-contiguous form only accounted for 5.48% in the learner corpus, which was smaller than the proportion in the reference corpus (12.00%). To put it another way, this finding suggested that Chinese EFL learners tended to overuse the contiguous form of separable PVs in their writing. Compared with the reference corpus, the chi-square test result also indicated the

significant overuse of the contiguous form in the Chinese learner corpus ($\chi^2 = 121.73$, $p < 0.001$)[26].

Figure 8.1 shows that Chinese EFL learners tended to overuse the contiguous form of separable PVs. In order to gain more details about the use of separable PVs in the learner corpus, Table 8.3 shows the number of separable PVs found in the two corpora and provides information about the frequency of the non-contiguous form.

Table 8.3 The non-contiguous form of separable PVs in each group: L1 users vs Chinese EFL learners

	Reference corpus			Learner corpus		
	#	Freq.	%	#	Freq.	%
Total	**55**	**720**	**12.00**	**40**	**223**	**5.48**
out	13	185	8.08	11	57	5.99
up	11	160	8.20	10	104	5.79
back	6	97	39.59	6	26	32.50
down	7	83	27.76	4	10	10.42
in	4	60	25.86	2	3	2.17
off	5	67	28.76	3	16	6.20
on	4	40	12.54	3	6	1.06
over	2	16	6.69	0	0	0
others	3	12	6.22	1	1	0.62

(*Note:* # means the number of separable PVs in the group.)

As Table 8.3 shows, there were 40 separable PVs in the learner corpus. Compared with the number found in the reference corpus (55 PVs), 14 separable PVs whose non-contiguous forms were absent from the learner corpus. For example, in the 'over' group, both *TAKE over* and *TURN over* were separable PVs, and there were 16 concordance lines in the reference showing the non-contiguous forms of these two PVs. The following two concordance lines showed the use of the non-contiguous form of *TAKE over* and

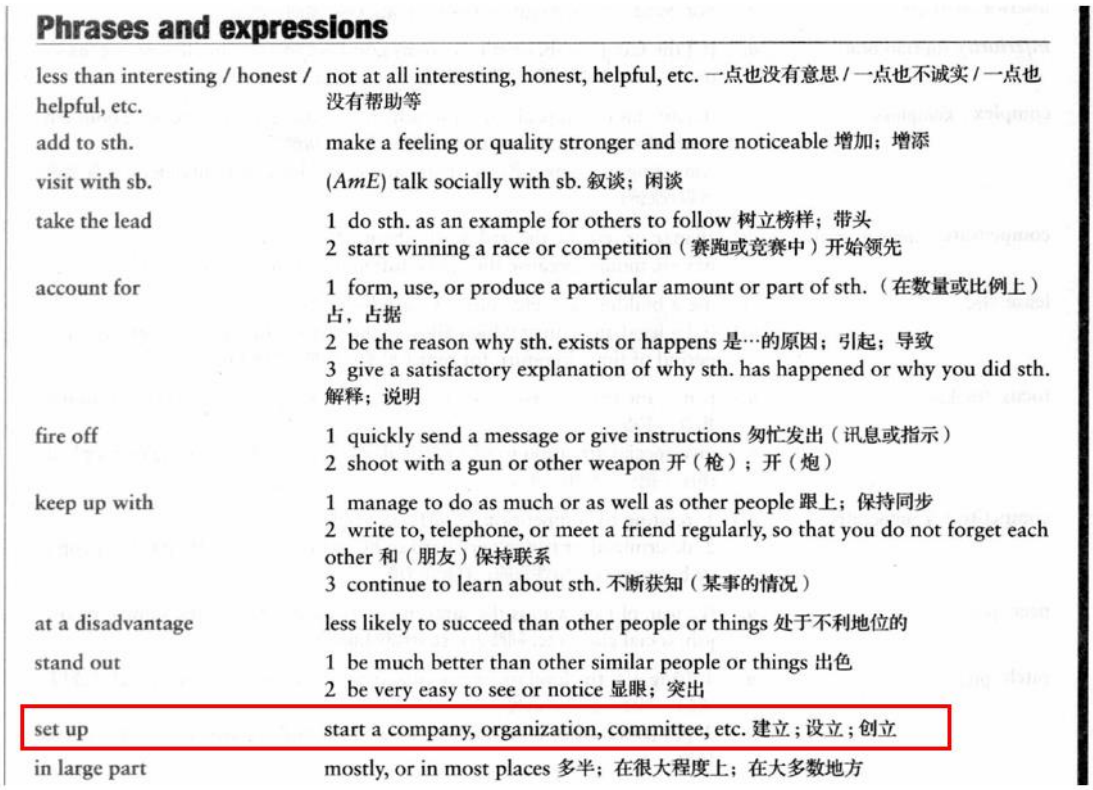
[26] In the reference corpus, there were 5280 occurrences showing the use of the contiguous form and 720 occurrences showing the non-contiguous form. In the Chinese learner corpus, however, the observed frequency was 3850 for the contiguous form and 223 for the non-contiguous form.

TURN over in L1 users' writing.

> By the time London Transport ***took them over*** in 1933, all but two had been so modified. (BNC: CBK)

> Sheff Wed next, who might be a bit tricky and are sure to ***turn some team over*** soon. (BNC: J1G)

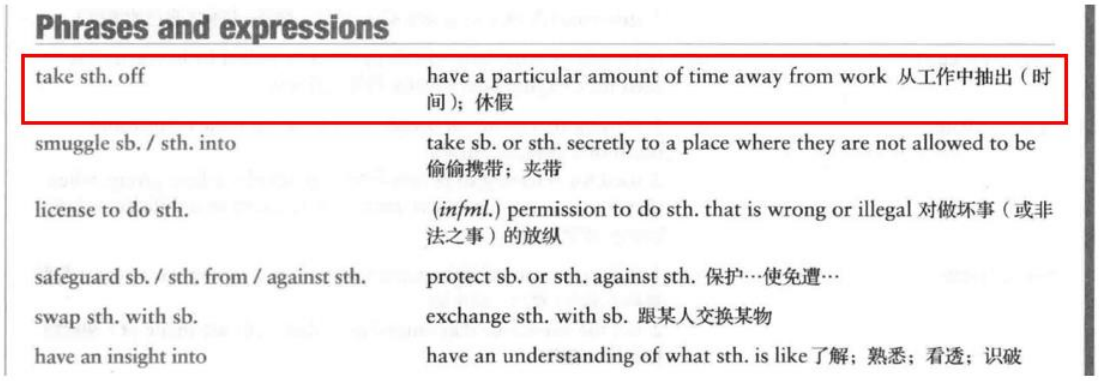
In learners' writing, however, there was no corpus evidence showing that Chinese learners were capable to employ the non-contiguous forms of *TAKE over* and *TURN over*. This suggested that learners have difficulty using the non-contiguous form of these two PVs. In addition to *TAKE over* and *TURN over*, the non-contiguous forms of PVs in other groups (e.g. *SET up, PUT out, PUT off*, etc.) were also absent from the learner corpus. One plausible cause for the absence could be related to the instructional influence. In *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015), Chinese EFL learners' course book of college English (see Section 4.4.2 for information about this course book), there was no example demonstrating that *SET up* can be used in a non-contiguous form (see Figure 8.2). The instruction of other separable PVs in this course book, such as *PUT out* and *PUT off*, can be found in Appendix 4.



Phrases and expressions

less than interesting / honest / helpful, etc.	not at all interesting, honest, helpful, etc. 一点也没有意思 / 一点也不诚实 / 一点也没有帮助等
add to sth.	make a feeling or quality stronger and more noticeable 增加；增添
visit with sb.	(*AmE*) talk socially with sb. 叙谈；闲谈
take the lead	1 do sth. as an example for others to follow 树立榜样；带头 2 start winning a race or competition（赛跑或竞赛中）开始领先
account for	1 form, use, or produce a particular amount or part of sth.（在数量或比例上）占，占据 2 be the reason why sth. exists or happens 是…的原因；引起；导致 3 give a satisfactory explanation of why sth. has happened or why you did sth. 解释；说明
fire off	1 quickly send a message or give instructions 匆忙发出（讯息或指示） 2 shoot with a gun or other weapon 开（枪）；开（炮）
keep up with	1 manage to do as much or as well as other people 跟上；保持同步 2 write to, telephone, or meet a friend regularly, so that you do not forget each other 和（朋友）保持联系 3 continue to learn about sth. 不断获知（某事的情况）
at a disadvantage	less likely to succeed than other people or things 处于不利地位的
stand out	1 be much better than other similar people or things 出色 2 be very easy to see or notice 显眼；突出
set up	start a company, organization, committee, etc. 建立；设立；创立
in large part	mostly, or in most places 多半；在很大程度上；在大多数地方

Figure 8.2 The instruction of separable PVs in *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015): *SET up*



Phrases and expressions

take sth. off	have a particular amount of time away from work 从工作中抽出（时间）；休假
smuggle sb. / sth. into	take sb. or sth. secretly to a place where they are not allowed to be 偷偷携带；夹带
license to do sth.	(*infml.*) permission to do sth. that is wrong or illegal 对做坏事（或非法之事）的放纵
safeguard sb. / sth. from / against sth.	protect sb. or sth. against sth. 保护…使免遭…
swap sth. with sb.	exchange sth. with sb. 跟某人交换某物
have an insight into	have an understanding of what sth. is like 了解；熟悉；看透；识破

Figure 8.3 The instruction of separable PVs in *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015): *TAKE off*

It should be noted that the learners' course book did provide guidance on the non-contiguous form of separable PVs. As Figure 8.3 shows, the course book gave clear instruction that *TAKE off* was a separable PV and demonstrated the non-contiguous use of this PVs. Corpus evidence from the Chinese learner corpus also indicated that learners were able to use the non-contiguous form of *TAKE off* in their writing when they were instructed to do so (see Figure 8.4).

Employees can not constantly **take time off** from work to learn; equally employers can not
I will **take two years off** in order to take care of my baby.
everyone in China is allowed to **take a few days off** spending time with

Figure 8.4 The non-contiguous form of *TAKE off* in the Chinese learner corpus

As a result, the findings showed that Chinese EFL learners significantly underused the non-contiguous form of PVs in their writing. Although learners were able to use the non-contiguous form of some PVs, they used this form less frequently than L1 users. The underuse seemed to be associated with the lack of explicit instruction in course books. While some separable PVs were clearly instructed in the course book (e.g. *TAKE off* in Figure 8.3), others were not (e.g. *SET up* in Figure 8.2). Chinese learners preferred the contiguous form of PVs and may feel uncertain about whether the verb and the particle in *SET up* should be separated. They need more exposure to separable PVs, especially those being absent from the learner corpus (e.g. *SET up, TAKE over, TURN over,* etc.). However, this argument is based on limited examples and should be considered tentative until further evidence is available to support a broader generalization.

8.1.3 The use of three-word PVs in the Chinese learner corpus

In Gardner and Davis's (2015) list of frequent PVs, they did not take PVs with three components into account. The current study includes three-word PVs because they are also units of meaning and have similar functions with individual verbs (see Section 2.1.2). Unfortunately, there was no list of frequent three-word PVs from the BNC or BoE corpora available for research at the time. This section firstly identifies frequent PVs with three components in the reference corpus, and then investigates how the use of three-word PVs differs in the Chinese learner corpus and the reference corpus.

As mentioned previously in Section 5.3, the two semantic criteria (i.e. meaning non-compositionality and meaning extension of the component) are applicable to three-word PVs. Using the two semantic criteria, *LOOK forward to* in the first corpus example below can be classified as a three-word PV because it has a non-compositional meaning. However, *GO back to* in the second example cannot be labelled as a three-word PV because its meaning can be predicted by the three components and none of these components carry a metaphorical or figurative meaning in this combination.

I ***look forward to*** receiving your information in the near future. (BNC: HJ8)

Leapor, on the other hand, could leave her job and ***go back to*** her father. (BNC: AN4)

Based on the semantic criteria, I investigated the most frequent three-word PVs and Table 8.4 provides an inventory of twenty frequent three-word PVs in the reference corpus.

Table 8.4 Twenty frequent PVs with three components in the reference corpus

	PV	Freq.	Freq. per million		PV	Freq.	Freq. per million
1	LOOK forward to	138	32.29	11	RUN out of	18	4.21
2	TURN out to	88	20.59	12	STAND up to	18	4.21
3	MAKE up of	50	11.70	13	GO along with	17	3.98
4	GET out of	42	9.83	14	PUT up with	15	3.51
5	DATE back to	34	7.96	15	GET on with	14	3.28
6	bound up with	27	6.32	16	MAKE up for	11	2.57
7	COME up with	26	6.08	17	END up with	10	2.34
8	LIVE up to	24	5.62	18	FACE up to	10	2.34
9	CATCH up with	18	4.21	19	GET away with	10	2.34
10	KEEP up with	18	4.21	20	LOOK down on	10	2.34
Total frequency: 598							

In this inventory, the most frequent three-word PVs was *LOOK forward to*, occurring 138 times (32.29 times per million words) in the reference corpus. The verbal

components in each frequent three-word PV were lemmatised, which was consistent with the approach used in the investigation of two-word PVs. It should be noted that in this inventory one three-word PV, *bound up with*, did not have various forms. When the verb *BIND* was followed by *up* and *with*, the three-word PV only had one form, *bound up with*, and this past participle form occurred 27 times in the reference corpus. As a result, this PV was not shown in its lemma form in Table 8.4.

As Table 8.4 shows, the twenty frequent three-word PVs occurred 598 times (139.93 per million words) in the reference corpus. In the Chinese learner corpus, the occurrences of these 20 PVs were 942 times (326.91 per million words). The normalised frequency indicated that Chinese EFL learners used three-word PVs more frequently than L1 users. This difference between the two corpora was statistically significant ($\chi^2 = 263.93$, $p < 0.001$). Compared with two-word PVs, Chinese learners seemed to be more familiar with three-word PVs and tended to overuse three-word PVs in their writing.

When I examined the use of each three-word PV in the Chinese learner corpus, the results showed that some of them were not frequently used by learners. There were four PVs that occurred fewer than ten times in the learner corpus, including *bound up with*, *END up with, GET away with*, and *STAND up to*. In other words, the high distribution in learners' writing was made up by fewer three-word PVs. Table 8.5 provides information about the frequency of sixteen three-word PVs in the learner corpus. The minimum frequency shown in Table 8.5 was 10 times. The frequency of the twenty frequent PVs in the learner corpus can be found in Appendix 2.

Table 8.5 The frequency of sixteen three-word PVs in the Chinese learner corpus

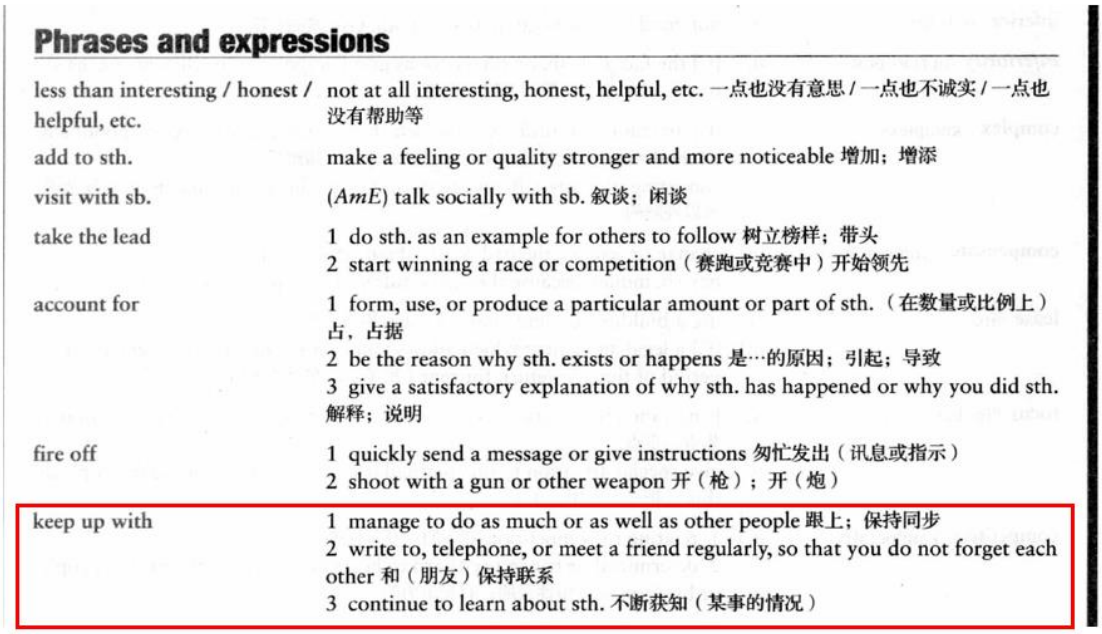
	PV	Freq.	Freq. per million		PV	Freq.	Freq. per million
1	LOOK forward to	295	102.34	9	FACE up to	25	8.67
2	KEEP up with	115	39.90	10	GET on with	24	8.33
3	CATCH up with	113	39.20	11	LIVE up to	20	6.94
4	COME up with	87	30.18	12	PUT up with	18	6.24
5	GET out of	49	17.00	13	DATE back to	16	5.55
6	LOOK down (up)on	49	17.00	14	MAKE up for	16	5.55
7	TURN out to	43	14.92	15	RUN out of	12	4.16
8	MAKE up of	35	12.14	16	GO along with	10	3.47

As Table 8.5 shows, Chinese learners frequently used *LOOK forward to* in their writing. The normalised frequency of this three-word PV in the learner corpus was 102.34 per million words, which was even higher than that in the reference corpus (32.29 per million words). This high distribution of this PV was probably due to the inclusion of empathy texts (i.e. emails and letters) in the learner corpus. The examination of this PV in the learner corpus showed that learners frequently used it at the end of emails and letters. For example,

We are ***looking forward to*** your early reply. (ID: TECCL01867)

I'm ***looking forward to*** your letter. (ID: TECCL09819)

Section 8.2.2 will revisit *LOOK forward to* and examine whether Chinese learners used it in a different way from L1 users based on more corpus examples.



Phrases and expressions

less than interesting / honest / helpful, etc.	not at all interesting, honest, helpful, etc. 一点也没有意思 / 一点也不诚实 / 一点也没有帮助等
add to sth.	make a feeling or quality stronger and more noticeable 增加；增添
visit with sb.	(*AmE*) talk socially with sb. 叙谈；闲谈
take the lead	1 do sth. as an example for others to follow 树立榜样；带头 2 start winning a race or competition（赛跑或竞赛中）开始领先
account for	1 form, use, or produce a particular amount or part of sth.（在数量或比例上）占，占据 2 be the reason why sth. exists or happens 是…的原因；引起；导致 3 give a satisfactory explanation of why sth. has happened or why you did sth. 解释；说明
fire off	1 quickly send a message or give instructions 匆忙发出（讯息或指示） 2 shoot with a gun or other weapon 开（枪）；开（炮）
keep up with	1 manage to do as much or as well as other people 跟上；保持同步 2 write to, telephone, or meet a friend regularly, so that you do not forget each other 和（朋友）保持联系 3 continue to learn about sth. 不断获知（某事的情况）

Figure 8.5 The instruction of three-word PVs in *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015): *KEEP up with*

In addition to *LOOK forward to*, Table 8.5 also indicates that learners frequently used *KEEP up with*, *CATCH up with* and *COME up with* in their writing. Learners' familiarity with these PVs seemed to be influenced by the instruction they received. Figure 8.5 shows the instruction of *KEEP up with* in the learners' course book of college English and the instruction for *CATCH up with* and *COME up with* can be found in Appendix 4. As Figure 8.5 demonstrates, Chinese learners had clear guidance on the three different meanings of *KEEP up with.* However, the course book did not provide explicit instruction on those less frequent three-word PVs, such as *bound up with, END up with,* and *STAND up to*. As a result, it can be argued that the frequency of three-word PVs in learners' writing was related to the instruction they received.

A closer examination of the use of three-word PVs in the learner corpus showed that Chinese EFL learners were inclined to choose *LOOK down upon* rather than *LOOK down on* in their writing. The frequency of *LOOK down upon* (37 occurrences) was three times

higher than *LOOK down on* (12 occurrences). However, the choice between these two PVs was completely different in L1 users' writing. In the reference corpus, *LOOK down on* occurred 10 times, but there was no corpus evidence showing the occurrence of *LOOK down upon*. As a result, Chinese EFL learners tended to be more familiar with *LOOK down upon*, whereas L1 users used *LOOK down on* more frequently.

Besides *LOOK down upon* and *LOOK down on*, the results also showed that Chinese learners and L1 users had different preferences for *GET on with* and *GET along with*. As Table 8.4 shows, *GET on with* was a frequent three-word PV which occurred 14 times in the reference corpus. *GET along with*, however, only occurred in one concordance line in the reference corpus even though it shares a similar meaning with *GET on with*. L1 users tended to choose *GET on with* rather than *GET along with* in written texts. In the Chinese learner corpus, the examination of the two PVs showed that *GET along with* was more frequent than *GET on with*. There were 220 concordance lines containing the use of *GET along with* in learners' writing, yet only 24 for *GET on with.* The substantial difference indicated that Chinese learners gave priority to *GET along with* and tended to avoid using *GET on with*.

Based on the analysis of *LOOK down on* and *GET on with* in the two corpora, it can be argued that Chinese EFL learners and L1 users had different preferences when they encountered a pair of synonymous PVs. The results showed that Chinese learners tended to be more familiar with those that were less frequent in the reference corpus (i.e. *LOOK down upon* and *GET along with*). They did not use *LOOK down on* and *GET on with* frequently in written texts. It seemed that learners were uncertain whether *LOOK down on* and *GET on with* worked in the same way as their synonymous PVs, *LOOK*

down upon and *GET along with*. Since quantitative data cannot tell the reason why Chinese learners preferred the less frequent option over the more frequent one, I examined the essay topics and the course book in order to find whether learners' habitual choice was influenced by the instruction in these materials. However, it seemed that learners' preference was probably influenced by their previous knowledge of *LOOK down upon* and *GET along with*, as the materials did not provide clear instruction on them (see Appendix 4). Learners may already have repetitive exposure to the two PVs, and they were more confident to use them in their writing.

8.1.4 Summary

In this section, I have examined the use of two-word and three-word PVs in the Chinese learner corpus and the reference corpus. Table 8.6 summarises the frequency of both two-word and three-word PVs in the two corpora. Overall, PVs were less frequent in the learner corpus than that in the reference corpus. The results showed that two-word PVs were significantly underused by Chinese EFL learners, while some three-word PVs, such as *LOOK forward to, KEEP up with, COME up with*, tended to be overused by learners.

Table 8.6 An overview of frequent PVs in the reference corpus and the Chinese learner corpus

	Reference corpus	Learner corpus
Freq. per million words	2370.83	2086.43
- two-word PVs	2241.43	1759.63
- three-word PVs	139.93	328.81

In the Chinese learner corpus, the frequency of two-word PVs was lower than that in the reference corpus. The data supports the findings from previous studies that EFL learner have difficulty using PVs, especially those whose L1 does not contain PVs (Dagut and Laufer, 1985; Waibel, 2007; Gilquin, 2015; Sung, 2020). Chinese learners

had difficulty using many PVs, especially those in the 'down', 'in' and 'over' groups (see Table 8.2). They probably need more exposure to two-word PVs and clear guidance on how to use these PVs in written texts.

It was also evident that Chinese EFL learners had difficulty using separable PVs. Compared with the number of separable PVs in the reference corpus, 14 separable PVs (such as *TAKE over, TURN over, SET up,* etc.) were absent from the learner corpus. Meanwhile, the non-contiguous form of separable PVs was significantly less frequent in learners' writing. The findings agree with Gilquin (2023) who found that ESL/EFL learners had trouble with the non-contiguous form. French ESL/EFL learners preferred the verb + particle + object structure instead of the verb + object + particle structure (Gilquin, 2023). In addition, the examination of separable PVs in the course book showed that Chinese learners' difficulties of the non-contiguous form was probably associated with the lack of explicit instruction in the course book. The instruction of *SET up* in the course book (see Figure 8.2), for example, did not provide clear guidance that this PVs allows the verb and the particle to be separated by the object. However, the course book gave clear instruction that *TAKE off* is a separable PV (see Figure 8.3). Corpus evidence showed that learners may be able to use the non-contiguous form of *TAKE off* more effectively when they received explicit instruction. However, this interpretation was based on the observed data from one course book should be explored further to determine whether it is a generalisable result.

In terms of three-word PVs, the frequency in the Chinese learner corpus was higher than that in the reference corpus. Based on the examination of 20 three-word PVs in the learner corpus, learners tended to be familiar with PVs such as *LOOK forward to*, *KEEP*

up with, COME up with. The frequent occurrences may be associated with the explicit instruction learners received from the course book. However, learners did not use *bound up with*, *END up with, GET away with*, and *STAND up to* frequently. Moreover, Chinese EFL learners and L1 users had different preferences when they encountered a pair of synonymous three-word PVs. For example, while *GET on with* was more frequent in L1 users' writing, learners preferred to use *GET along with*. A potential reason for learners' preference for *GET along with* and *LOOK down upon* could be their previous knowledge of these two PVs.

8.2 The use of specific PVs in the Chinese learner corpus

8.2.1 The use of *LOOK up* in the Chinese learner corpus

In Section 8.1, I have examined the overall performance of using PVs in the learner corpus. It seemed that PVs in the 'up' group may be overused by Chinese EFL learners because of the high frequency of this group in the learner corpus (see Table 8.2). In the 'up' group, *LOOK up* was one of the most frequent PVs in learners' writing, occurring in 217 concordance lines. As a result, this section focuses on *LOOK up* and compares the use of this PV in the reference corpus and the learner corpus.

8.2.2 Frequency and form

Table 8.7 provides information about the frequency of *LOOK up* in the two corpora. As Table 8.7 shows, *LOOK up* occurred 217 times (75.28 per million words) in the learner corpus. However, in the reference corpus, it only occurred 46 times (10.76 per million words). The disproportionate frequency of *LOOK up* in the learner corpus suggested that learners may overuse this PV in their writing.

Table 8.7 Frequency of *LOOK up* in the reference corpus and the Chinese learner corpus

	Reference corpus	Learner corpus
Frequency	46	217
Freq. per million	10.76	75.28

One possible cause for the disproportionate frequency could be the small size of the Chinese learner corpus. The small size suggested that the learner corpus represented a limited range of language use, which may result in an abnormal frequency of specific linguistic features per million words (Gablasove, Brezina and McEnery, 2017). In addition, the biased selection of written tasks in the learner corpus may also lead to a disproportionate result. There was one written task that asked students to discuss the advantages and disadvantages of E-dictionaries (see topic 22 in Appendix 3). This topic seemed to have a strong association with the high frequency of *LOOK up* in the learner corpus (see Section 8.2.1.2 for more detail).

While the frequency of *LOOK up* in the learner corpus seemed to be disproportionate, it should be cautious to interpret this quantitative data. The two inherent problems in learner corpus research (i.e. the small size of the learner corpus and the biased selection of learner' written texts) may affect the validity of the quantitative data. In order to enhance the understanding of how Chinese learners use *LOOK up* in written texts, the current study also analysed qualitative data, including the meanings and grammatical patterns associated with this PV. The results will be outlined in the following sections.

Since *LOOK up* is a separable PV in English, I also compared the occurrences of the contiguous and non-contiguous forms of this PVs in the two corpora. The results of the comparison are outlined in Table 8.8.

Table 8.8 Different forms of *LOOK up* in the reference corpus and the Chinese learner corpus

Forms	Reference corpus		Learner corpus	
	Freq.	%	Freq.	%
LOOK up	42	87.50	191	88.02
LOOK + up	6	12.50	25	11.52
LOOK ++* up	0	0	1	0.46
Total	48		217	

(*Note:* + represents an arbitrary token, and * represents an optional token)

The data in Table 8.8 suggested that both L1 users and Chinese EFL learners used the contiguous form of *LOOK up* more frequently. In the two corpora, the occurrences of the non-contiguous form decreased as the distance between *LOOK* and *up* increased. In the reference corpus, the non-contiguous form occurred 6 times, accounting for 12.50% of the total occurrences of this combination. In the learner corpus, there were more occurrences (26 occurrences) showing the use of the non-contiguous form, accounting for 11.98% of the total occurrences. Moreover, as Table 8.8 shows, L1 users and Chinese learners tended to insert one word between *LOOK* and *up*. In both corpora, there were only a small number of instances in which the distance between the two components exceeded two words. The learner corpus contained one example where *LOOK* and *up* were separated by three words. Here is the example:

> It's faster and easier to ***look*** the new words ***up*** in e-dictionary than in traditional dictionary. (ID: WARG3492)

In the Chinese learner corpus, the normalised frequency of *LOOK up* was higher than that in the reference corpus. This disproportionately high frequency suggested that learners tended to overuse this PV in written texts. The results also showed that Chinese EFL learners were able to use the non-contiguous form of *LOOK up*, though both L1 users and Chinese learners were inclined to use the contiguous form in their writings.

However, the quantitative data might be unreliable due to the limitations of learner corpus research (i.e. the small size of the learner corpus and the restricted selection of learners' written text in the corpus). As a result, a closer examination of the qualitative data would be necessary in order to make a stronger argument that learners have difficulty using *LOOK up* in their writing.

8.2.2.1 Meanings of *LOOK up* in the Chinese learner corpus

In Section 6.1, the examination of *LOOK up* in the BoE corpus showed that this combination had five distinctive meanings. Among the five meanings, the literal meaning *LOOK up*$_1$ and the non-compositional meaning *LOOK up*$_4$ tended to be more frequent than the other three meanings. This section begins with a question: can Chinese learners recognize and use the five different meanings of *LOOK up* in their writings? To answer this question, I investigated the frequency and proportion of each meaning in the reference corpus and then compared with the results gained from the learner corpus. Table 8.9 shows the frequency of each meaning in the reference corpus.

Table 8.9 Distinctive meanings of *LOOK up* in the reference corpus

	Meanings	Frequency	Percentage (%)
LOOK up$_1$	to use or raise eyes physically to see someone or something	23	47.92
LOOK up$_2$	to improve or become better	4	8.33
LOOK up$_3$	to admire or respect someone	3	6.25
LOOK up$_4$	to consult or find a piece of information from a source	17	35.42
LOOK up$_5$	to visit somebody	1	2.08

As Table 8.9 shows, the literal meaning of *LOOK up* (i.e. *LOOK up*$_1$) occurred 23 times, which was the most frequent meaning in the reference corpus, accounting for nearly half of the total occurrences (47.92%). It was consistent with the finding I gained in the

BoE corpus, suggesting that L1 users tended to use *LOOK up* in a literal way. In terms of the four PV usages, *LOOK up*$_4$ was more frequent than the other three meanings, making up 35.42% of the occurrences in the reference corpus (17 occurrences). In the reference corpus, *LOOK up*$_5$ was the least popular meaning, with only one concordance line demonstrating the association between the meaning of *LOOK up* and visiting. Moreover, the frequency of *LOOK up*$_2$ (4 occurrences) was slightly higher than the frequency of *LOOK up*$_3$ (3 occurrences).

Table 8.10 Distinctive meanings of *LOOK up* in the Chinese learner corpus

	Meanings	Frequency	Percentage (%)
LOOK up$_1$	to use or raise eyes physically to see someone or something	7	3.23
LOOK up$_2$	to improve or become better	0	0
LOOK up$_3$	to admire or respect someone	8	3.69
LOOK up$_4$	To consult or find a piece of information from a source	201	92.63
LOOK up$_5$	to visit somebody	1	0.46

Table 8.10 provides information about the frequency of the five meanings in the learner corpus. As Table 8.10 shows, *LOOK up*$_4$ was the most frequent meaning in learners' writing, occurring 201 concordance lines. In addition to *LOOK up*$_4$, learners were able to use *LOOK up*$_1$ and *LOOK up*$_3$ in their writing, but the occurrences of these two meanings (7 occurrences and 8 occurrences respectively) were significantly fewer than the occurrences of *LOOK up*$_1$.

Figure 8.6 shows the percentage of each meaning of *LOOK up* in the two corpora. In the learner corpus, the percentage of each meaning was different from that in the reference corpus. Chinese EFL learners frequently used *LOOK up*$_4$ in their writing (92.63%), but they were less likely to use the other four meanings in written texts.

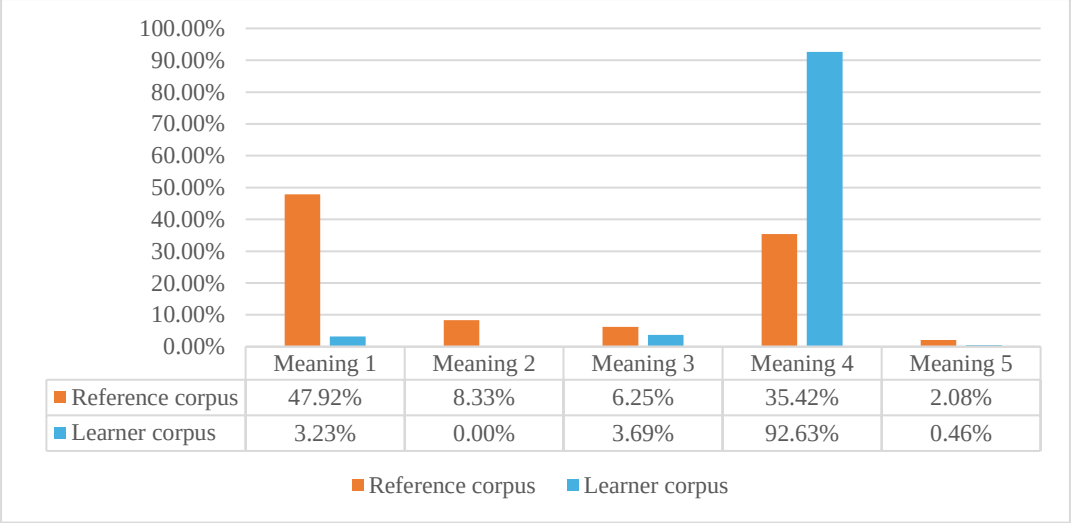


Figure 8.6 The proportion of each meaning of *LOOK up* in the reference corpus and the Chinese learner corpus

Moreover, there was no corpus instance showing the occurrence of *LOOK up*$_2$ in the learner corpus. This indicated that *LOOK up*$_2$ was a meaning that Chinese EFL learners were not familiar with. They had very restricted exposure to this specific meaning. In the reference corpus, the frequent collocation for *LOOK up*$_2$ was 'things + *LOOK up*' (see Section 6.1.1). However, Chinese learners did not use this collocation, and they tended to express this idea in a different way. Figure 8.7 exhibits five concordance lines in which learners employed an alternative to 'things + *LOOK up*'. In these examples, learners regarded expressions such as 'things are getting better' or 'things will be better' as alternatives to 'things + *LOOK up*'. These expressions were semantically congruent with 'things + *LOOK up*', indicating a positive prediction about the future.

1 what I'm inclined to. So **things** are going well for you, right?
2 Though I am still shy, **things** are getting much better. I'm delighted to see
3 getting a good grade. I hope **things** will be better for you soon.
4 With such efforts, I believe **things** will become much better.
5 own state of mind. I believe **things** will get better soon.

Figure 8.7 The alternative use for 'things + *LOOK up*' in the Chinese learner corpus

Based on the examination of *LOOK up* in the Chinese learner corpus, the frequency of each meaning was associated with the level of familiarity. The results showed that Chinese EFL learners were particularly familiar with *LOOK up*$_4$, the non-compositional

meaning of this PV. In the learner corpus, the frequency of *LOOK up*$_4$ was even higher than that in the reference corpus. This was probably due to the selection of written texts in the learner corpus (see topic 22 in Appendix 3). Besides *LOOK up*$_4$, it can be argued that Chinese EFL learners were able to use *LOOK up*$_1$ and *LOOK up*$_3$ although the results suggested that the frequency of these two meanings was not as frequent as *LOOK up*$_4$. However, learners tended to be unfamiliar with *LOOK up*$_2$ and *LOOK up*$_5$, because the two meanings were extremely less frequent in the learner corpus. This suggested that learners had very limited exposure to *LOOK up*$_2$ and *LOOK up*$_5$ when they learned this PV.

8.2.2.2 Grammatical patterns associated with *LOOK up* in the Chinese learner corpus

In the reference corpus, the frequent pattern associated with *LOOK up*$_4$ was ***LOOK up*** **n** ***in*** **n** and Figure 8.8 provides some examples to demonstrate this pattern. Within this pattern, the first noun referred to the information being consulted (e.g. 'the meaning of manipulate'), whereas the second noun indicated the source from which the information can be consulted (e.g. 'the Oxford English Dictionary', 'a book', etc.). Lines 7 to 9 show that *LOOK up*$_4$ is separable. In other words, the pattern can also be written as ***LOOK*** **n** ***up in*** **n**, with the verb and the particle being separated by the first noun. L1 users tended to use ***LOOK*** **n** ***up in*** **n** when *LOOK* and *up* were separated by pronouns (see lines 7 to 9).

6 One woman had **looked up** the meaning of manipulate **in** the Oxford English Dictionary.
7 if you can go and **look** it **up in** a book? I just thought of it as
8 What's prostitute?" I **look** them **up in** in the big dictionary
9 tell each other our dreams and **look** them **up in** these absurd books,

Figure 8.8 The frequent grammatical pattern associated with *LOOK up*$_4$ in the reference corpus

On the other hand, the examination of *LOOK up*4 in the Chinese learner corpus showed that learners misused this PV in some cases. Examples of misuse can be found in Figure 8.9.

10 about the word. So when we **look up** a traditional dictionary we must use
11 and we want to **look up** an E-dictionary because it bring us either speed
12 us to carry with. It makes us **look up** the dictionary easily wherever we are
13 easily to form a habit of **looking up** the E-Dictionary.
14 need some information you can **look up** the Internet more urgent need.

Figure 8.9 Misuse of *LOOK up*4 in the Chinese learner corpus

Table 8.11 The frequency and proportion of the misuse of *LOOK up*4 in the learner corpus

	Frequency	Percentage (%)
Misuse of *LOOK up*4	63	31.34
Proper use	138	68.66
Total	201	

When Chinese EFL learners employed *LOOK up*4 in their writing, they tended to use this PV in a problematic pattern, ***LOOK up* n**. In this problematic pattern, the noun following the PV indicated the source where the information can be consulted. To be more specific, learners were confused about the function of the noun that followed *LOOK up*4, and the preposition *in* was missing in learners' use. As Table 8.11 shows, there were 63 concordance lines showing the problematic use of *LOOK up*4, which is more than one-third of the total frequency in the learner corpus (31.34%). Chinese learners' ignorance of prepositions was probably due to the interference of their native language. Chinese does not have many prepositions compared to English, and learners' knowledge of L1 may habitually and unconsciously influence their use of L2 (Li, Ren and Zhao, 2016). In Chinese, when people consult the meaning of a word in the dictionary, they usually say *cha zidian* (查 字典), with *cha* (i.e. *LOOK up*) being the verb and *zidian* (i.e. dictionary) being the object. There was no preposition needed in the Chinese expression.

As a result, when Chinese learners switched from their L1 to L2, they were likely to produce an English expression without any preposition. The ignorance of prepositions in learners' use can be regarded as an example of negative transfer.

The examination of *LOOK up*$_4$ in the learner corpus also showed that Chinese EFL learners have difficulty differentiating between *LOOK up*$_4$ and *LOOK for*. As the concordance lines in Figure 8.10 illustrate, *for* seemed to be redundant in the three sentences. Learners made mistakes by keeping both *up* and *for* in one sentence. For example, in line 15, '*look up for* some information of the history' was incorrect. It should be acceptable if learners use '*look up* some information of the history' or '*look for* some information of the history'. This problem also revealed that learners have difficulty using prepositions.

15 the sentence a second time and **looked up** for some information of the history
16 For the moment you **look up** for the very word you want to know
17 from abroad it is impractical to **look up** for the words in the dictionary

Figure 8.10 Confusion between *LOOK up*$_4$ and *LOOK for* in the Chinese learner corpus

In the BoE corpus, there were some examples showing that *up* and *for* can co-exist after *LOOK* in one sentence (see lines 18 to 20 in Figure 8.11). When *up* and *for* co-occurred, the meaning of '*LOOK up for*' tended to be associated with *LOOK up*$_2$. Unfortunately, learners did not use this specific meaning in their writing.

18 Life is **looking up for** English football in the lower divisions
19 Things are finally **looking up for** this town.
20 Things were also **looking up for** impotent men in the United States

Figure 8.11 The concordance lines of '*LOOK up for*' in the BoE corpus

As a result, while the quantitative data suggested that Chinese EFL learners were familiar with the non-compositional meaning of *LOOK up* (i.e. *LOOK up*$_4$), the examination of the grammatical pattern showed that learners sometimes misused this

PV in written texts. They were confused about the function of the noun following *LOOK up*$_4$ and made mistakes by ignoring the preposition in the pattern ***LOOK up*** **n** ***in*** **n**. The qualitative analysis also showed learners' confusion about *LOOK up*$_4$ and *LOOK for*. Learners' difficulties were potentially associated with the lack of prepositions in their L1. In order to reduce errors, Chinese EFL learners need more exposure to *LOOK up*, especially the grammatical patterns associated with *LOOK up*$_4$.

8.2.3 The use of *LOOK forward to* in the Chinese learner corpus

As mentioned in Section 8.1.3, *LOOK forward to* was the most frequent three-word PV in the two corpora. It occurred 138 times (30.76 per million words) in the reference corpus and 295 times (102.34 per million words) in the learner corpus. Based on the corpus data, it can be implied that Chinese EFL learners were familiar with this three-word PV. This section examines how Chinese learners used this three-word PV in their writing, with particular focuses on the verb forms and patterns associated with it.

8.2.3.1 Different tense forms of *LOOK forward to* in the Chinese learner corpus

When L1 users and Chinese EFL learners used *LOOK forward to* in their writing, the data showed that they had different preferences for the choice of verb forms. Figure 8.12 demonstrates the proportion of different verb forms in the two corpora.

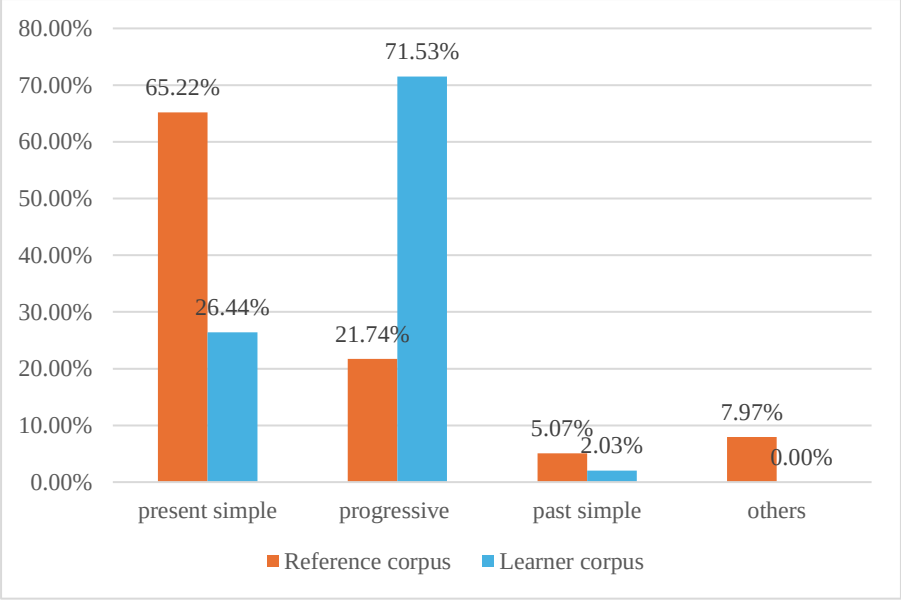


Figure 8.12 Different verb forms of *LOOK forward to* in the reference corpus and the Chinese learner corpus

As Figure 8.12 shows, both L1 users and Chinese EFL learners were inclined to use the present simple form or the progressive form of *LOOK forward to* in written discourse, especially in letters and emails. However, the present simple form (i.e. *look forward to* or *looks forward to*) was more frequent in the reference corpus, whereas the progressive form (i.e. *looking forward to*) was more frequent in the learner corpus.

In the reference corpus, more than half (65.22%) of the occurrences showed the use of the present simple form of *LOOK forward to*. The progressive form was less frequent than the present simple, accounting for 21.74% of the total occurrences. The 'others' category in Figure 8.12 includes other forms such as the present perfect form (i.e. have *looked forward to*) and the form with modal verbs. L1 users sometimes used *LOOK forward to* with modal verbs such as *can, may, could,* etc. (see examples in Figure 8.13). Unfortunately, this modality usage was absent from the learner corpus.

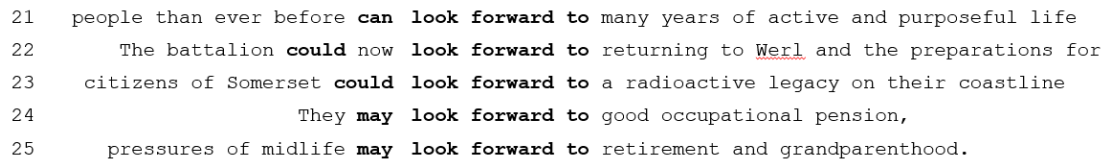


Figure 8.13 *LOOK forward to* with modal verbs in the reference corpus

In the Chinese learner corpus, the progressive form of *LOOK forward to* was substantially more frequent than other verb forms, accounting for 71.53% of the occurrences. Learners also used the present simple form in some cases (25.61%), but other forms, including the past simple, were extremely rare in their writing. The forms of *LOOK forward to* were less various in the learner corpus. Learners may have more exposure to the progressive form of this PV, and they sticked to using this particular form in written texts. Moreover, as the most frequent form in the learner corpus, it was noteworthy that *looking forward to* often occurred at the beginning of the sentence, without following the subject and the auxiliary verb '*be*'. For example,

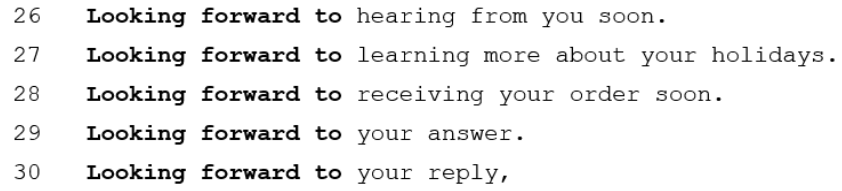


Figure 8.14 The frequent usage of *looking forward to* in the Chinese learner corpus

Surprisingly, this specific usage of *looking forward to* was absent from the reference corpus. The evidence from the reference corpus suggested that L1 users would not omit the subject and the auxiliary verb when they employed *looking forward to* in written texts. This may be related to the stylistic differences across registers, as the corpus evidence from the BoE corpus showed that this usage often occurred in spoken texts (see Figure 8.15).

31 I don't really mind. **Looking forward to** it so much.
32 Saturday morning? Okay. **Looking forward to** it. Bye-bye.
33 coming to the end of the journey? **Looking forward to** a hot bath?

Figure 8.15 *Looking forward to* (without the subject and the auxiliary verb) in spoken texts in the BoE corpus.

Figure 8.15 provides corpus evidence that the ellipsis of the subject and the auxiliary verb before *looking forward to* can be found in spoken texts. The ellipsis of the subject and the auxiliary verb tended to result in an informal and colloquial tone. In written language, however, learners are expected to produce complete sentences with an explicit use of subjects and auxiliary verbs.

Since Chinese EFL learners often used *looking forward to* with the ellipsis of the subject and auxiliary verb, it can be argued that they cannot distinguish the stylistic differences between spoken and written language. As a result, language teachers should give clear instructions on the use of *LOOK forward to* in both written and spoken discourse. Learners should be aware that the ellipsis of the subject and auxiliary verb is rare in written discourse. It is in fact an indicator of the spoken language.

8.2.3.2 Frequent patterns associated with *LOOK forward to* in the Chinese learner corpus

In addition to the use of different tense forms, the examination of *LOOK forward to* in the reference corpus and the learner corpus showed that L1 users and Chinese EFL learners frequently used this PV in two patterns, **V n** and **V -ing**. Table 8.12 below provides information about the frequency of these two patterns in the two corpora.

Table 8.12 Grammatical patterns associated with *LOOK forward to* in the reference corpus and the Chinese learner corpus

	Reference corpus		Learner corpus	
	Frequency	Percentage (%)	Frequency	Percentage (%)
V -ing	94	68.12	125	38.11
V n	42	30.43	181	55.18

As Table 8.12 shows, when L1 users and Chinese EFL learners employed *LOOK forward to* in written texts, they have different preferences for the two patterns. In the reference corpus, the occurrences of '*LOOK forward to v-ing*' accounted for more than two-thirds of the total occurrences (68.12%), which was higher than the percentage of '*LOOK forward to n*' (30.43%). In other words, the **V -ing** pattern was more frequent than the **V n** pattern in L1 users' writing. However, Chinese EFL learners tended to make an opposite choice. In the learner corpus, more than half of the concordance lines (55.18%) demonstrated the use of the **V n** pattern, whereas only 38.11% showed the use of the **V -ing** pattern. This result suggested that learners prefer '*LOOK forward to n*' over '*LOOK forward to v-ing*'.

In learners' writing, when *LOOK forward to* was followed by nouns, this PV seemed to have a communicative function. Figure 8.16 provides some examples illustrating this specific usage observed in the learner corpus.

```
34          it is convenient for you? Looking forward to your answer. Yours,
35      be very suitable for you. I'm looking forward to your arrival.
36     morning, May the third. We are looking forward to your attendance. Please inform
37   relationship for a long time. We look forward to your reply by May 30th,
38    her to express herself and I am looking forward to your response. Yours sincerely
```

Figure 8.16 The use of '*LOOK forward to n*' in the Chinese learner corpus

Concordance lines in Figure 8.16 demonstrated that Chinese learners tended to use nouns such as 'answer', 'arrival' and 'reply' to follow *LOOK forward to*. The examination of these lines showed that '*LOOK forward to n*' indicated an interaction

between the sender and the recipient of the letter. For example, when '*looking forward to* your answer' was used in a letter (line 34), it suggested that the sender was willing to discuss a question and requested a reply from the recipient.

In the reference corpus, Figure 8.17 exhibits some concordance lines that contain the use of '*LOOK forward to n*'. Lines 39 and 40 illustrated that L1 users also showed their willingness and requested a communication with the recipient by employing '*look forward to* a reply' and '*look forward to* their comments'. However, this communicative meaning tended to be less evident in lines 41 to 45. For instance, the expression '*looking forward to* an increased interest' in line 42 implied an expectation of a higher rate of interest. L1 users expressed their expectations rather than their willingness to engage in a communication in lines 41 to 45.

39	your urgent attention. I	**look forward to**	a reply by return of post.
40	to the appropriate committee and	**look forward to**	their comments. Yours sincerely
41	crew are smiling as they clear up,	**looking forward to**	a well-deserved shower
42	supported by all our staff. "We are	**looking forward to**	an increased interest in this
43	on the lake. We were	**looking forward to**	some exercise of rowing,
44	contribution to this. We	**look forward to**	the creation of the DSDC.
45	will appreciate how much they	**look forward to**	the day that they can return

Figure 8.17 The use of '*LOOK forward to n*' in the reference corpus

The genre of texts may affect the communicative function of *LOOK forward to*. In the learner corpus, both '*LOOK forward to v-ing*' and '*LOOK forward to n*' were frequently used in letters or emails. Learners rarely used this PV in other genres, such as argumentative essays. In the reference corpus, '*LOOK forward to v-ing*' was more frequent in letters and emails, whereas '*LOOK forward to n*' may occurred in other text genres. In letters and emails, corpus evidence also showed that L1 users conveyed their willingness through the use of '*LOOK forward to v-ing*'. Figure 8.18 provides some corpus examples to show the communicative meaning indicated by '*LOOK forward to*

v-ing'.

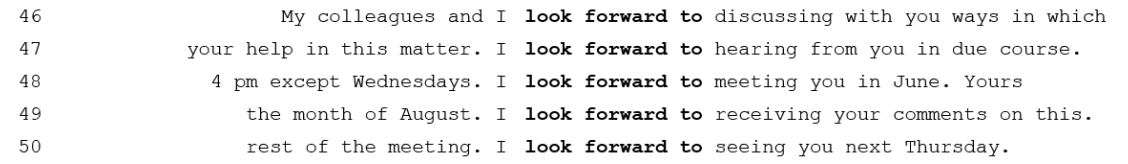


Figure 8.18 The communicative meaning of '*LOOK forward to v-ing*' in the reference corpus

Based on the analysis, it can be argued that Chinese learners used *LOOK forward to* in a different way from L1 users. The results suggested that learners tended to use the progressive form whereas L1 users used the present simple form more frequently. When learners used *looking forward to*, they often omitted the subject and the auxiliary verb. However, such ellipsis indicated an informal and colloquial tone, which was inconsistent with the stylistic feature of the written language. The ellipsis of the subject and the auxiliary verb before *looking forward to* suggested that Chinese EFL learners were confused about stylistic differences between the spoken and written language. Moreover, while Chinese EFL learners were able to use the two frequent grammatical patterns associated with *LOOK forward to* (i.e. '*LOOK forward to n*' and '*LOOK forward to v-ing*'), they tended to be more familiar with '*LOOK forward to n*'. The results showed that learners habitually employed expressions such as '*looking forward to* your answer' and '*looking forward to* your response' (lines 34 and 38 in Figure 8.16) to imply their willingness to engage in a communication and request a response from the recipient. This was different from L1 users' choice. In L1 users' writing, '*LOOK forward to n*' often implied an expectation (e.g. '*look forward to* an increased interest'), whereas '*LOOK forward to v-ing*' was more likely to convey their willingness to communicate.

8.3 The choice between PVs and one-word verbs

Previous studies found that EFL learners tended to choose one-word verbs rather than PVs in spoken and written discourse (Liao and Fukuya, 2004; Siyanova and Schmitt, 2007). However, Chapter 7 has outlined the difficulties of relacing PVs with synonymous one-word verbs (see Sections 7.1 and 7.2). In this section, I will firstly investigate Chinese EFL learners' verb choices between PVs and one-word verbs in written texts, and then examine their performance when replacing PVs with synonymous one-word verbs.

8.3.1 L1 users' and Chinese EFL learners' choices between PVs and one-word verbs

In written texts, multi-word verbs tend to be less frequent than their one-word alternatives. This tendency has been confirmed by Siyanova and Schmitt (2007) who studied 26 multi- and one-word verb pairs in the written sub-corpus of BNC and the International Corpus of Learner English (ICLE). According to their study, the results for L1 users and EFL learners were consistent. In the two corpora, most multi-word verbs (18 out of 26) were less frequent than their corresponding one-word equivalents (Siyanova and Schmitt, 2007). As an important category of multi-word verbs, it can be speculated that PVs are also less frequent than their one-word equivalents in written texts.

In the current study, I investigated the frequency of 10 PVs and their one-word equivalents in the Chinese learner corpus. Table 8.12 below lists the 10 PVs, and the corresponding one-word verbs selected in this study. Out of the 10 PVs, 8 were two-

word PVs that were frequently used in learners' writing, and the remaining 2 were selected from the list of 20 frequent three-word PVs identified in Section 8.1.3. These 10 PVs were intended to represent the range of English PVs that learners may be exposed to. Corpus evidence showed that they were frequent; some were polysemous (e.g. *GO up, TURN down*) while others had a consistent meaning (e.g. *GET back*, *HOLD on*). It should be noted that these synonymous one-word verbs could not be perfect alternatives to PVs, especially polysemous PVs. For example,

> He just ***turns down*** the volume on his computer. (BoE: NC4__041205)
>
> We would certainly not ***turn down*** a request for such a meeting to explain our policy. (BoE: NB8__040224)

The meaning of *TURN down* in '*turns down* the volume' is different from the meaning in '*turn down* a request for a meeting'. The former is synonymous with 'decrease', while the latter is related to 'refuse'. As a result, this study has a limitation that the chosen one-word alternatives may not be perfect replacements and learners may have individual preferences for one-word replacements.

Table 8.13 Normalised frequency of the 10 verb pairs in the reference corpus and the Chinese learner corpus

Pairs		Reference corpus		Learner corpus	
PV	OWV	PV	OWV	PV	OWV
GET back	return	16.72	200.81	8.32	90.12
GO up	rise	23.63	92.05	12.20	108.14
HOLD on	wait	34.77	94.72	30.78	126.72
LOOK for	search	51.95	36.04	85.69	100.96
PUT off	postpone	8.02	24.52	2.77	14.14
SET up	establish	122.81	265.89	56.84	113.10
TURN down	decrease	12.14	16.72	11.09	64.88
WORK out	solve	37.44	39.45	26.06	376.28
CATCH up with	reach	3.34	174.07	31.89	144.74
GET out of	avoid	21.17	130.61	15.25	153.62
	Total	331.99	1074.87	329.93	1292.70

(Note: OWV is the abbreviation for 'one-word verb')

To investigate whether Chinese EFL learners avoid using PVs and prefer one-word verbs over PVs, Table 8.13 provides information about the frequency of these 10 PVs and their one-word alternatives in the reference corpus and the learner corpus. As Table 8.13 shows, PVs were less frequent than one-word verbs in the two corpora. There was no statistically significant difference in the use of PVs and one-word verbs in the two corpora ($\chi^2 = 4.70$, $p > 0.05$). Therefore, it can be argued that PVs are generally less frequent in written texts. Both L1 users and Chinese EFL learners tend to use one-word verbs in their writing.

Table 8.13 is an overview of the use of PVs and their one-word alternatives in the two corpora. Although PVs were less frequent in the two corpora, this did not necessarily mean that both L1 users and Chinese EFL learners avoided using PVs and tended to replace PVs with one-word verbs. For example, in the reference corpus, *LOOK for* occurred 51.95 times per million words, which was more frequent than its one-word alternative, *search* (36.04 per million words). Figure 8.19 below shows the different choices between PVs and one-word verbs in the two corpora. In Figure 8.19, there are three types of differences. The label 'PV > OWV' indicates that the PV is more frequent than its one-word alternative, whereas the label 'PV < OWV' indicates that the PV is less frequent than its one-word alternative. In addition, the label 'PV ≈ OWV' indicates that the normalised frequency of the PV and the one-word alternative are approximately equal (the difference is less than 5 per million words).

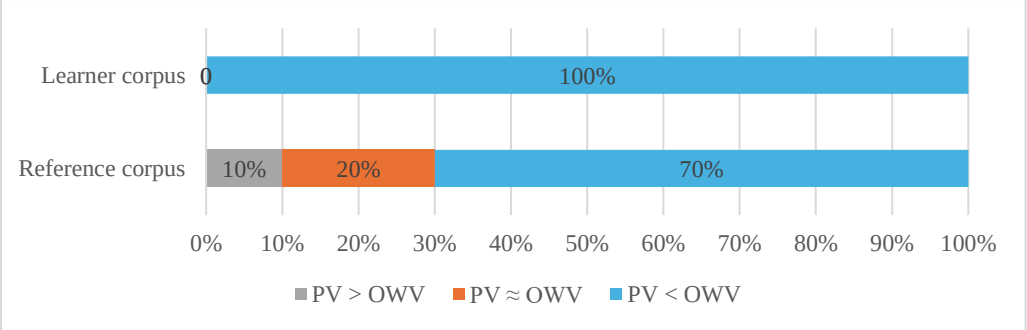


Figure 8.19 Different choices between PVs and one-word verbs in the reference corpus and the Chinese learner corpus

As Figure 8.19 shows, in the learner corpus, the normalised frequency of all the 10 PVs were lower than their one-word alternatives. This result suggested that Chinese learners had a stronger tendency to choose one-word verbs rather than PVs. They tended to avoid using PVs when they had a one-word option. However, in the reference corpus, as mentioned earlier, the frequency of *LOOK for* was even higher than its one-word alternative. In addition, there were 2 out of 10 PVs whose frequencies were approximately equal to their one-word alternatives. According to the data shown in Table 8.12, these two PVs were *TURN down* and *WORK out*. As a result, Chinese EFL learners always chose one-word verbs and avoided using PVs in their writing. L1 users made different choices in some cases. L1 users did not have a strong preference for one-word words when they encountered *LOOK for*, *TURN down* and *WORK out*.

Findings from the current study and Siyanova and Schmitt's (2007) research provide converging evidence that PVs were less frequent than their one-word alternatives in written texts. In the current study, Chinese EFL learners had a preference for one-word verbs. The results suggested that learners tended to avoid using PVs and replaced PVs with one-word verbs. However, L1 users tended to make different choices when they dealt with specific PVs such as *LOOK for*, *TURN down* and *WORK out*.

8.3.2 Chinese learners' performance when they replace PVs with one-word verbs

As mentioned in Section 8.3.1, English PVs usually have multiple meanings, and it is hard to find a perfect replacement for a polysemous PV. Findings from Section 7.1 also supported this argument. While *search* seems to have a synonymous meaning with *LOOK for*, the examination of *LOOK for* and *search* showed that they were not interchangeable because they were used in a different way from the perspective of pattern grammar (see Section 7.1 for more details). Due to grammatical differences, learners should be cautious when they replace PVs with one-word alternatives. Given the grammatical differences between PVs and one-word verbs, I examined the use of *LOOK for* and *search* in the Chinese learner corpus and investigated whether Chinese EFL learners are able to recognize the grammatical differences between them.

In the learner corpus, *LOOK for* occurred 247 times (85.69 per million words), which was less frequent than *search* occurring 291 times (100.96 per million words). Before looking into the grammatical use of the PV and the one-word verb, I have examined whether Chinese learners regarded *search* as the one-word alternative to *LOOK for*. Table 8.14 provides evidence to show that *search* was regarded as the one-word alternative to *LOOK for* in the Chinese learner corpus. As Table 8.14 shows, the PV and the one-word verb shared three nominal collocates, including *job, information, word*.

Table 8.14 Frequent nominal collocates of look for and search in the learner corpus

LOOK for		**SEARCH**	
	Frequency		Frequency
job	50	information	54
information	26	internet	51
money	8	job	12
word	7	word	7
dictionary	6		

(Note: a. all the nominal collocates are lemmatized; b. the minimum frequency is 5)

In terms of the grammatical use, Section 7.1 has shown that the patterns where *LOOK for* and *search* frequently appear are different. In the BoE corpus, *LOOK for* frequently appears in the **V n *in* n** pattern, whereas *search* cannot be used in this pattern. For example,

> Ms Colwill is ***looking for*** a job **in** the city and plans to commute by train. (BoE: NA1__040912)
>
> *Ms Colwill is ***searching*** a job **in** the city and plans to commute by train.

As the example shows, *looking for* cannot be substituted by *searching* directly. A possible replacement could be '*search for* a job in the city'. As mentioned in Section 7.1, the frequent patterns in which *search* is used are **V *for* n** or **V n *for* n** (see Section 7.1). To explore whether Chinese EFL learners used *search* in a different way from *LOOK for*, Table 8.15 provides information about the use of this one-word verb in the Chinese learner corpus.

Table 8.15 The use of *search* in the Chinese learner corpus

	Frequency	Percentage (%)
A. '*SEARCH* for n'	103	40
B. '*SEARCH* n for n'	34	13
C. Neither A nor B (e.g. '*SEARCH* on the Internet', etc.)	35	14
D. Incorrect use of *search* (e.g. '*SEARCH* some information on the Internet, etc.)	85	33

The use of *search* was manually categorised into four different groups. As Table 8.15 shows, '*SEARCH* for n' was more popular than other usages in the learner corpus. While learners preferred to use '*SEARCH* for n', they seemed to be less familiar with '*SEARCH* n for n' because it only occurred 34 times in the learner corpus. Besides the two frequent patterns associated with this verb, Table 8.15 also shows a disappointing result. About

one-third of the occurrences demonstrated the incorrect use of *search* in the learner corpus. This result suggested that Chinese EFL learners have difficulty using *search* in their writing. Here are some corpus examples showing the incorrect use.

```
51          that people want to search a better job by attending the university
52   you can do other things. To search a work in your spare time.
53    dictionary when we want to search an unfamiliar word, and more
54  most of us are interested in searching information on the Internet instead of
55       We can use computers to search some information and watch
```

Figure 8.20 The incorrect use of *search* in the Chinese learner corpus

As an intransitive verb, it is impossible for *search* to be followed by the object directly. Therefore, expressions such as '*search* a better job' (line 51 in Figure 8.20) and '*search* an unfamiliar word' (line 53) were incorrect. Based on this result, when learners replaced *LOOK for* with *search*, it can be speculated that learners may ignore the grammatical differences between them. The findings from *LOOK for* and *search* in the learner corpus suggested that learners may have difficulty recognising the grammatical differences when they replaced PVs with one-word verbs. They tended to focus on the semantic equivalence but ignore the grammatical differences between PVs and one-word verbs.

To sum up, based on the investigation of the choice between PVs and one-word verbs, it can be argued that Chinese EFL learners had a strong tendency to use one-word verbs and avoided using PVs in written discourse. However, corpus evidence suggested learners were not capable of recognizing the grammatical differences between PVs and one-word verbs. When they replaced PVs with one-word verbs, they would give priority to the semantic equivalence rather than the grammatical differences between PVs and their one-word alternatives.

8.4 Summary

To sum up, this chapter has compared with the use of frequent PVs in the Chinese learner corpus and the reference corpus. The results showed that two-word PVs in the learner corpus were significantly less frequent than that in the reference corpus, whereas three-word PVs were more frequent in the learner corpus. Chinese EFL learners also underused the non-contiguous form of frequent PVs, suggesting that they had difficulty using separable two-word PVs. The qualitative analysis of *LOOK up* showed that learners also struggled to use this PV in written texts. They were confused about the grammatical pattern associated with this PV and made mistakes by ignoring the preposition after it (see Section 8.2.1).

Similar to ESL/EFL learners from other L1 backgrounds, Chinese EFL learners also avoided using PVs and preferred one-word verbs over PVs. Moreover, when Chinese learners replaced PVs with one-word verbs, they usually focused on the semantic equivalence between PVs and one-word alternatives but ignored the grammatical differences between them. Because of the ignorance, learners often made mistakes when replacing PVs with one-word verbs.

Chapter 9 Conclusion

The final chapter of the thesis will summarize the key findings and present conclusions of the current research. In Section 9.1, I will revisit the research questions and answer each of them based on the findings that I have covered in chapters 6 to 8. Section 9.2 discusses pedagogical implications of the research, with a particular focus on what teachers and course book writers need to do differently in accordance with the conclusions presented in Section 9.1. Section 9.3 outlines the limitations as well as highlighting some strengths of the research. Finally, I will propose some areas for future investigation.

9.1 Discussion

The study focuses on English PVs, examining the lexico-grammatical patterns associated with specific PVs and comparing the use of PVs in the reference corpus and the Chinese learner corpus. Before answering each research question and discussing the significance of the findings, it should be noted that the current study used semantic criteria to identify and classify verb + particle combinations in the corpus. This was different from the traditional method that identified PVs using syntactic tests (cf. Bolinger, 1971; Biber et al., 1999; Fraser, 1976) and differentiated between PVs and other verb + particle combinations based on the grammatical function of the particle (i.e. adverbial or prepositional). It was argued in the current study that traditional syntactic tests of PVs were unreliable and the grammatical labels on the non-verbal component tended to be arbitrary because they were largely dependent on linguists' intuition (Section 2.2). For example, from the traditional perspective, *COME across* in the following sentence may not be classified as a PV because this combination does not

allow the movement of the particle.

He ***came across*** a professional actor. (BoE: BA_Nm012157)

However, PVs stand out from other multi-word combinations because of their specialised meanings, ranging in compositionality from transparent verb + particle combinations to relatively opaque combinations (Moon, 1997). In the case of *COME across*, although it does not allow the movement of the particle, the meaning of this combination is non-compositional. As a result, this study did not use syntactic tests to identify PVs but proposed two semantic criteria to classify verb + particle combinations in the corpus. The first semantic criterion was concerned with whether the meaning of verb + particle combinations can be predicted by the individual components (i.e. whether the meaning of the combination was compositional or not). Using this criterion, *COME across* in the given example was classified as a PV (see Section 5.1.1). However, Sinclair (1996) noted that meaning compositionality cannot be the formal criterion because the individual components may have other relationships with the meaning of PVs. Therefore, the study proposed a second semantic criterion, assessing whether the verbal or non-verbal component had an extended meaning when it was combined with the other. Using the second semantic criteria, *LOOK into* was classified as a PV in the following corpus example because the non-verbal component *into* was used in a figurative way.

We're ***looking into*** the situation. (BoE: BU_fM042762)

In addition to two-word PVs, the two semantic criteria were also applicable to three-word PVs, such as *PUT up with* and *LOOK down on* (see Section 5.3 for corpus examples). Furthermore, the study also noted the variation in the use of verb + particle combinations (Section 5.2). A verb + particle combination could have both PV and non-PV usages. Using semantic criteria to classify verb + particle combinations had to be

tied to the context in which the combination appeared. One cannot claim that *COME across* is a PV without showing the context where this combination appears.

9.1.1 How do L1 users use specific PVs in terms of three aspects: collocation, patterns, and the progressive form?

The various usages also reflected the complexity of English PVs. Given the complexity, the first research question of the current study was concerned with the lexico-grammatical patterns associated with specific PVs in the BoE corpus and examined how collocation and grammatical patterns helped to disambiguate the multiple meanings of PVs. The results of *LOOK up* showed that it had five distinctive meanings, and each meaning co-occurred with different nominal collocates (see Section 6.1.3). Corpus evidence suggested that each meaning of *LOOK up* was restricted by specific nominal collocates. More precisely, when nouns such as *eyes* or *sky* co-occurred with *LOOK up*, it was the literal meaning (i.e. *LOOK up*$_1$) that made sense in the context. When the pattern 'things + *LOOK up*' (i.e. *LOOK up* follows the noun *things* in the co-text) was examined, the data indicated that *LOOK up*$_2$ was exclusive in this pattern. Other meaning varieties of *LOOK up* did not occur within this pattern. This finding is consistent with Sinclair's (1991) idiom principle in that the co-selection of lexical items constitutes a single choice. Semantic congruency can be observed between the meaning of PVs and the specific choice of collocates in the co-text. For example, *LOOK up*$_1$ frequently co-occurred with *sky* and *stars*, whereas *LOOK up*$_2$ was less likely to co-occur with these nouns. Moreover, as a polysemous PV, the meanings of *LOOK up* were used in different forms. *LOOK up*$_2$ and *LOOK up*$_3$ were exclusively used in the contiguous form. The other two meanings, *LOOK up*$_4$ and *LOOK up*$_5$, had both the contiguous and non-contiguous forms, but L1 users showed a strong tendency to insert an intervening item between the

verb and the particle when they employed these two.

In the BoE corpus, it was found that some PVs are more likely to be used in the progressive form, such as *LOOK up*$_2$, *LOOK for*, and *LOOK forward to*. Previous studies argued that the progressive form can suggest users' negative attitude, especially when the form was combined with *always* or other adverbials of this kind (Römer, 2005; Lindley, 2020). The data of *LOOK for* supports this argument. When *looking for* was combined with *always*, as the following example shows, it connoted a negative meaning, suggesting the speaker's complaint about the excessive spending.

> He is always ***looking for*** ways to obtain money, but spends it like water when he has it. (BoE: BB_Ff90_715)

Moreover, the findings also showed that the adverbial *always* was indispensable for the negative connotation of *LOOK for*. Without *always*, the emotional meaning of *looking for* seemed to be neutral in the following concordance line.

> The important objective is to try and find out what the customer is really ***looking for***. (BoE: NC5__031227)

In contrast to *LOOK for*, when *LOOK up*$_2$ was used in the progressive form, it was less likely to capture users' negative attitude. Given the metaphorical meaning of *up*, it may not be surprising that *looking up* in the following sentence conveyed a positive connotation.

> Things were definitely ***looking up***. Paul Warren was coming back to work today. (BoE: BU_jM012174)

This progressive use of *LOOK up*$_2$ provides convincing evidence for Mindt's (1997) argument. While the progressive form was more likely to express negative emotions, Mindt (1997) argued that it also expressed positive emotions in some cases (cf. Mindt,

1997, p.237). Corpus evidence of *LOOK up*$_2$ supports this argument that L1 users also express their positive emotions via the progressive form.

In terms of *looking forward to*, when this three-word PV had *always* in the co-text, corpus evidence showed that both positive and negative connotations can be observed. For instance,

> She was <u>always</u> ***looking forward to*** what she would be doing when she got out of school and preparing for that day (BoE: NU4__021123)
>
> He says that they were <u>always</u> ***looking forward to*** doing their next movie together and never got the chance. (BoE: NB3__040124)

In the first example, *looking forward to* suggested the positive expectation of the future, whereas in the second example, it was associated with a negative feeling of disappointment.

Finally, I also investigated the grammatical patterns associated with specific PVs and their potential one-word synonyms in the BoE corpus. Hunston and Francis (2000) argued that patterns can be used to differentiate meanings of a polysemous word (e.g. *afraid of* and *afraid that* in Section 3.2.2). The results of this study agree with this argument. The meaning of *LOOK up* in the pattern ***LOOK up*** **n** ***in*** **n** was different from the meaning in ***LOOK up to*** **n**. For example,

> I have to keep on ***looking up*** words ***in*** the dictionary or thesaurus. (BoE: BB_Yf022010)
>
> Milton, as a small boy, had ***looked up to*** his father as small boys usually do. (BoE: BA_Qf022206)

The present findings, and those of Hunston and Francis (2000), provide evidence to show that patterns help to disambiguate polysemous items including individual words and multi-word combinations. Another important argument in Hunston and Francis's (2000) study is that words from the same semantic group tend to be used in the same pattern. Based on this argument, this study hypothesised that a PV and its potential one-word synonyms may have a similar pattern. However, the results do not support this hypothesis. Not all PVs share the same pattern with their synonymous one-word verbs. For example, *LOOK for* and *search* were a pair of near synonyms, but they were used in different grammatical patterns (see Section 7.1). Because of the grammatical differences, *LOOK for* cannot be replaced with *search*. For instance,

We're going to ***look for*** a cottage in the Highlands. (BoE: BB_cF93_893)

*We're going to ***search*** a cottage in the Highlands.

LOOK for frequently appeared in the **V n *in* n** pattern, whereas *search* did not work properly in this pattern. Although the findings do not support Hunston and Francis's argument of pattern and meaning, the distinctive patterns associated with *LOOK for* and *search* suggest that PVs cannot be directly substituted by one-word verbs.

Taking account of the results presented in Chapters 6 and 7, it can be argued that collocation and grammatical patterns help to clarify the differences between the meanings of polysemous PVs. There is semantic congruency between frequent collocates and the meaning of PVs. Each meaning is also used in specific pattern. The lexico-grammatical patterns play a crucial role in understanding the multiple meanings of PVs. Moreover, the answer of the first research question also suggests that PVs cannot be replaced with one-word equivalents because they are often used in different patterns. This point will be revisited when answering the third research question (see

Section 9.1.3).

9.1.2 How does the use of frequent PVs differ between the reference corpus and the Chinese learner corpus?

To answer the second research question, I investigated the use of frequent PVs in Chinese EFL learners' writing, including both two-word and three-word PVs, and compared with the use in L1 users' writing. The answer to this question helps to understand Chinese learners' difficulties of using PVs in written discourse and have significant implications for teaching and learning PVs.

9.1.2.1 The use of two-word PVs

Based on the data collected from learners' written texts, this study provided corpus evidence showing that Chinese EFL learners significantly underused two-word PVs in their writing. This result is consistent with the findings in previous studies that EFL learners whose native language does not contain the phrasal verb structure usually struggle with using English PVs (Waibel, 2007; Gilquin, 2015; Sung, 2020). For example, Sung (2020) observed a similar trend in a Korean-speaking learner corpus and identified twenty-four PVs that were significantly underused by Korean learners. Although the current study did not investigate the most underused PVs in the Chinese learner corpus, the results showed that Chinese learners found PVs in the 'down', 'in' and 'over' groups more challenging than PVs in other groups (see Section 8.1.1). This supports the findings in Sung's (2020) study because Korean learners also had greater difficulty using PVs such as *BREAK down*, *BRING in*, *TAKE over*, etc. Among the twenty-four underused PVs, Sung (2020) observed that Korean learners did not use PVs in the

'up' group frequently, such as *BRING up*, *SET up*, *COME up*. However, the results of the current study indicated that PVs in the 'up' group were more frequent than PVs in other groups (Section 8.1.1). Chinese learners seemed to be more familiar with PVs in the 'up' group, perhaps because they found some PVs more useful than others when writing essays. For example, *LOOK up* was highly sensitive to the essay topic and most instances of *LOOK up* were associated with one essay topic (see topic 22 in Appendix 3). As a result, while the absence of PVs in learners' L1 is a key factor in their underuse of PVs, the data in the Chinese learner corpus and the data from Sung's (2020) study shows that EFL learners from various L1 backgrounds face challenges with different sets of PVs.

The results from a closer examination of *LOOK up* suggested that Chinese EFL learners were not very proficient in using *LOOK up*, even though they frequently employed this combination in their writing. As mentioned in Section 6.1, *LOOK up* is a polysemous two-word PV with five distinctive meanings. In the reference corpus, the most frequent meaning of *LOOK up* was the literal meaning, referring to 'raising eyes in order to see something' (i.e. *LOOK up*$_1$). This finding was not surprising because previous studies also regarded *LOOK up*$_1$ as the most important meaning of *LOOK up* (e.g. Garnier and Schmitt, 2015; Liu and Myers, 2020). Garnier and Schmitt (2015) suggested that the most frequent meaning should be more important than other less frequent meanings. Given the importance of *LOOK up*$_1$, it was reasonable to hypothesise that Chinese EFL learners would use the literal meaning more frequently than other meanings. However, the results from the learner corpus did not support the hypothesis. The literal meaning was less frequent in the Chinese learner corpus, and Chinese EFL learners seemed to be more familiar with the non-compositional meaning, *LOOK up*$_4$ (i.e. to consult or find information from a source). Also, Chinese learners were able to use the metaphorical

meaning of *LOOK up* (i.e. *LOOK up*$_3$) in their writing. The results were surprising because they rejected the intuitive assumption that "literal meanings [of PVs] should be better known than figurative ones" (Garnier and Schmitt, 2016, p.31). Given the low frequency of *LOOK up*$_1$ and high frequency of *LOOK up*$_4$ and *LOOK up*$_3$ in the learner corpus, Chinese EFL learners seem to pay more attention to the idiomatic and figurative meanings of PVs rather than the literal and more frequent meaning. The findings of *LOOK up* in the current study do not support previous studies in that high frequency in the reference corpus does not necessarily lead to frequent use in the learner corpus.

In Section 8.2.1, I also analysed the use of *LOOK up*$_4$ in the Chinese learner corpus. The results of the qualitative analysis revealed that learners sometimes made mistakes when they employed *LOOK up*$_4$ in their writing. The following example demonstrates the typical problematic use of *LOOK up*$_4$ in the Chinese learner corpus.

> *When we ***look up*** the Chinese-English E-Dictionaries, there may be many words having the similar meaning. (ID: WARG0216)

In the reference corpus, however, the frequent grammatical pattern associated with *LOOK up*$_4$ is ***V*** **n** ***in*** **n**, as in the following concordance line:

> One woman had ***looked up*** the meaning of manipulate ***in*** the Oxford English Dictionary. (BNC: G0T)

The comparison between the two concordance lines showed that Chinese learners selected different nouns to follow *LOOK up*$_4$. As the example extracted from the reference corpus shows, the noun phrase, 'the meaning of manipulate', directly follows the PV, which is concerned with the item that is being focused on. The preposition phrase, 'in the Oxford English Dictionary', appears after the noun phrase in order to introduce the source of this piece of information. However, the problematic use found

in the learner corpus suggested that learners were confused about the semantic functions of the two nouns in the pattern ***LOOK up* n *in* n**. They made mistakes due to a lack of phraseological knowledge of *LOOK up*$_4$, producing the unnatural expression '*look up* the Chinese-English E-Dictionaries' and ignoring the preposition *in* in the pattern. In other words, Chinese learners were unaware of the lexico-grammatical pattern associated with *LOOK up*$_4$ and had difficulty using prepositions. This agrees with Li, Ren and Zhao (2016), who examined the grammatical mistakes in Chinese college students' writing. Due to the incongruency between learners' L1 and L2, Chinese learners often selected wrong prepositions unconsciously (Li, Ren and Zhao, 2016). For example, they found that learners were confused about 'pay attention on' and 'pay attention to' in some cases (Li, Ren and Zhao, 2016, p.26). The findings from the current study provide additional evidence to show learners' difficulty with prepositions. When Chinese learners used *LOOK up*$_4$ in their writing, they were unable to differentiate the semantic functions of the nouns and ignored the preposition in ***LOOK up* n *in* n**.

Another important finding was that Chinese EFL learners struggled to use the non-contiguous form of PVs in their writing. The current study showed that the number of separable PVs in the Chinese learner corpus (40 PVs) was fewer than the number observed in the reference corpus (55 PVs). Moreover, the frequency of the non-contiguous form was significantly lower than that in the reference corpus (see Section 8.1.2). In other words, Chinese EFL learners were inclined to use the contiguous form of separable PVs in their writing, with the object appearing directly after the particle (i.e. *look up the word* rather than *look the word up*). The results corroborate the findings from Gilquin (2015) who argued that French-speaking learners also preferred the 'verb + particle + object' structure (i.e. the contiguous form) over the 'verb + object + particle'

structure (i.e. the non-contiguous form). There is a semantic reason underlying the placement of the particle (Gries and Stefanowitsch, 2004; Gilquin, 2015). The non-contiguous form tends to be preferred with combinations whose meanings are compositional or partially compositional, such as *TURN off* and *TAKE out*[27] (Gries and Stefanowitsch, 2004, p.112). Gardner and Davis (2018) also found that the major grammatical structure for *TURN off* and *TAKE out* was 4-gram strings; that is, the verb and the particle were separated a two-word intervening item. However, in the Chinese learner corpus, learners kept using the contiguous form of *TURN off* and *TAKE out*, regardless of their meanings. For example,

> You also have to ***turn off*** the lights on time. (ID: WARG3527)
>
> When we arrive home, we ***take out*** our goods and throw away the bags unconsciously. (ID: WARG2144)

The underuse of the non-contiguous form seemed to be associated with the way that separable PVs were treated in learners' course books. Unfortunately, the findings from the chosen course book showed that Chinese EFL learners were not given explicit instruction that separable PVs had both contiguous and non-contiguous forms (see Appendix 4). In the course book, separable PVs, such as *SET up*, were typically treated as fixed phrases, without guidance on the use of the non-contiguous form. However, if clear guidance on the non-contiguous form was available in the course book, learners would be able to use this form. For example, the course book provides clear instruction that *TAKE off* can be used in a non-contiguous way, especially when it expresses 'having amount of time away from work' (see Table 8.3 in Section 8.1.2). Following the

[27] In Gries and Stefanowitsch's (2004) study, they used the label, 'non-idiomatic combinations', to refer to verb + particle combinations whose meanings are compositional or partially compositional. The particle in a non-idiomatic combination had a literal or metaphorical meaning (Gries and Stefanowitsch, 2004). Combinations whose meanings are non-compositional, were referred to as 'idiomatic combinations' by Gries and Stefanowitsch (2004).

instruction, learners were able to use the non-contiguous form of *TAKE off*. The following concordance line illustrates the non-contiguous form of *TAKE off* in the Chinese learner corpus.

Employees cannot constantly ***take*** time ***off*** from work to learn. (ID: WARG2172)

Consequently, it can be argued that the underuse of the non-contiguous form in the Chinese learner corpus is associated with the lack of explicit instruction. This is crucial to teaching PVs and I will return to this point in the pedagogical implications (Section 9.2).

9.1.2.2 The use of three-word PVs

To date the use of three-word PVs in learners' writing has not yet been extensively examined in previous studies. This study touched on this area and showed that the overall use of three-word PVs in Chinese learners' writing was more satisfactory than the use of two-word counterparts. Compared with L1 users, Chinese learners showed a strong tendency to use three-word PVs frequently in their writing. There seemed to be several reasons for the frequent use of three-word PVs. Firstly, compared to two-word PVs, the structure of three-word PVs is more fixed. Also, the meanings of three-word PVs are more consistent than two-word counterparts. Wei (2021) argued that PVs were usually taught as chunks in Chinese English classrooms and students tended to memorize the meaning of PVs mechanically. Although Wei (2021) did not provide evidence to support this argument, the instruction of three-word PVs in the course book could reflect the way of teaching PVs in Chinese context. Three-word PVs, such as *CATCH up with*, *COME up with*, and *KEEP up with*, were treated as fixed phrases in the course book (see Appendix 4). The fixed structure and the consistent meaning of three-word PVs would make them easier to memorise, increasing the opportunity of using in

written texts.

Secondly, Chinese learners seem to have particular preference for some less frequent three-word PVs. Learners' preference can be reflected by the high frequency of *GET along with* and *LOOK down upon* in the Chinese learner corpus. In the reference corpus, however, these two PVs were extremely rare because L1 users were more likely to use *GET on with* and *LOOK down on* in written texts (see Section 8.1.3). For example,

> Students could learn how to ***get along with*** others in the college. (ID: WARG1579)
>
> We try to ***get on with*** people, to be friendly with them and all we get is trouble. (BNC: B0N)

The two examples show that Chinese learners and L1 users have different preferences. Although both *GET along with* and *GET on with* have similar meanings in the two examples, learners prefer to choose *GET along with*, the one that is less frequent in the reference corpus. In addition to Chinese learners, Gilquin (2015) argued that French-speaking learners also had special preference which was distinctive from L1 users. In Gilquin's (2015) study, she regarded *GET along* and *GET on* as two-word PVs and found that French-speaking learners preferred *GET along*, whereas L1 users preferred *GET on*. Such distinction was probably due to the fact that *GET along* had a more transparent meaning than *GET on* (Gilquin, 2015). However, the verb *GET* is delexicalized in the two PVs, suggesting that the meaning of both *GET along with* and *GET on with* is not completely transparent. Therefore, in addition to semantic reasons, EFL learners' distinctive preference may be associated with other reasons.

In the current study, I argue that learners' preference for *GET along with* and *LOOK down*

upon is influenced by their previous knowledge of PVs. The distinctive preferences of L1 users and EFL learners are inconsistent with the vocabulary teaching strategy. Many linguists have acknowledged the crucial role of frequency in vocabulary acquisition (Nation, 2013; Siyanova-Chanturia and Webb, 2016). The frequency-based strategy is considered an effective strategy when it comes to vocabulary acquisition (Nation, 2013; Schmitt and Schmitt, 2014). Learners have more opportunities to encounter and acquire a word if it occurs frequently (Durrant et al., 2022; Milton and Hopwood, 2022). According to the frequency-based strategy, EFL learners should be more familiar with PVs that are frequently used by L1 users. However, the findings from the current study and Gilquin's (2015) study showed that EFL learners' habitual choice was not determined by frequency. Learners preferred less frequent yet semantically transparent PVs, such as *GET along with* and *LOOK down upon*, rather than those frequent PVs in the reference corpus (i.e. *GET on with* and *LOOK down on*). Learners' habitual choice is perhaps related to their previous knowledge of PVs. EFL learners tend to be more confident to use words they are familiar with because they can easily retrieve them from their previous knowledge and avoid making mistakes (Schmitt, 2000; Nation, 2013). Learners may have repeated exposure to *GET along with* and stored it in their knowledge of PVs. When they need an expression conveying 'having good friendship with someone', they tend to stick to the one they have already known (i.e. *GET along with*) rather than the one they are not familiar with (i.e. *GET on with*).

9.1.3 Do Chinese EFL learners replace PVs with one-word equivalents in their writing? How satisfactory are their replacements?

The third research question of this thesis focused on the use of PVs and their one-word equivalents in the Chinese learner corpus. The investigation into the frequency of ten

PVs and their corresponding one-word equivalents showed that PVs were less frequent than their one-word equivalents in the learner corpus. In other words, Chinese EFL learners tended to avoid PVs in their writing and preferred one-word verbs over PVs. The findings agree with Liao and Fukuya (2004) who also observed the avoidance behaviour of Chinese-speaking learners, especially learners at the intermediate level. Previous studies have concluded that EFL learners tend to replace PVs with one-word verbs (Dagut and Laufer, 1985; Liao and Fukuya, 2004; Siyanova and Schmitt, 2007) and learners usually regard the replacement as a "play-it-safe strategy" because one-word verbs may have more general and less idiomatic meanings than PVs (Hulstijn and Marchena, 1989, p.241). In the current study, Chinese learners also showed a strong tendency to replace PVs with one-word verbs. For example, in the Chinese learner corpus, while *LOOK for* and *search* had similar nominal collocates (e.g. *job*, *information*, *word*), the frequency of *LOOK for* was lower than *search* (see Section 8.3.2). This implied that learners regarded *LOOK for* and *search* as a pair of synonyms, but they tended to use the individual verb more frequently, probably because *search* has a clearer and more transparent meaning than *LOOK for*.

Based on the data gathered from a spoken learner corpus, Wei (2021) concluded that Chinese learners at lower proficiency levels (primary and intermediate levels) avoided using PVs in spoken discourse. The current study gathered data from a written learner corpus and showed that Chinese EFL learners also avoided using PVs and prefer one-word verbs over PVs in written discourse. The present findings, and those from Wei's (2021) study, provide corpus evidence to support the idea that Chinese EFL learners, especially those at an intermediate proficiency level, struggle with using PVs in written and oral communication. Because of their difficulty with PVs, they tend to avoid PVs

and prefer one-word verbs over PVs in both written and spoken discourse.

Apart from Chinese EFL learners' avoidance behaviour, a closer examination of *LOOK for* and *search* suggested that learners' performance was not satisfactory when they replaced *LOOK for* with *search*. Learners regarded *LOOK for* and *search* as a pair of synonyms and used them interchangeably in their writing. This gave rises to the problematic use such as the following:

> It is no doubt that internet is the best way to ***look for*** information. (ID: WARG3124)
>
> *Sometimes it's very convenient and fast to ***search*** the information from the Internet. (ID: WARG1671)

The two examples extracted from the learner corpus are almost identical, except for the choice of verbs preceding *information.* An examination of the second example showed that learners used *search* in a different way from L1 users. Learners' use of *search* in the second example was problematic, with *for* being absent from their production. In the BoE corpus, when *search* co-occurred with *information*, the most frequent pattern was **V *for* n**, as in the following example:

> 'To google' means to ***search for*** information on the Internet. (BoE: SU3__040514)

When answering the first research question, the results showed that *LOOK for* and *search* appeared in different grammatical patterns in the BoE corpus. In other words, *search* cannot be treated as the one-word synonym for *LOOK for* from the perspective of pattern grammar. However, the absence of *for* after *search* in learners' writing suggests that Chinese learners directly replace *LOOK for* with *search*. They seem to give priority to semantic equivalence when they replace PVs with one-word verbs, ignoring the

grammatical differences between the PV and the one-word verb.

Finally, previous studies concluded that PVs are more frequent in spoken discourse (Biber et al., 1999). Use of PVs in written texts may lead to an informal and colloquial tone (Siyanova and Schmitt, 2007). Waibel (2007) speculated that the reason of preferring one-word verbs over PVs may be associated with the stylistic feature of PVs and one-word verbs. Students tended to regard one-word verbs as more sophisticated words, making their writing more native-like and formal (Waibel, 2007). The data from the Chinese learner corpus supports this viewpoint that Chinese learners use more one-word verbs and avoid using PVs in their writing. However, it is not evident that Chinese learners are sensitive to the stylistic differences between written and spoken discourse. The data showed that the frequency of three-word PVs in the Chinese learner corpus was higher than the frequency in the reference corpus (see Section 8.1.3). Given the high frequency of three-word PVs in learners' writing, it can be argued that Chinese EFL learners are not sensitive to the stylistic features of PVs. The stylistic difference between PVs and one-word verbs may not be the primary reason for the avoidance of PVs. Based on the findings from the learner corpus, I argue that the main reason is Chinese learners' difficulty with PVs. While they are more familiar with one-word verbs, they also make mistakes when replacing PVs with one-word verbs, paying too much attention to the semantic equivalence but ignoring the grammatical difference between the PV and the one-word verbs.

9.2 Implications for the classrooms

The conclusions of the current study provide some implications for language teaching,

not only for the teaching of PVs itself, but also for the teaching of PVs through collocation and grammatical patterns. I will discuss the two aspects in the following two sections.

9.2.1 Teaching PVs

First of all, the answer to the second research question (Section 9.1.2) shows the underuse of two-word PVs in Chinese learners' writing. Chinese learners also use PVs in a different way from L1 users. They are less likely to use the contiguous form of PVs in their writing. Wei (2021) suggested that learners' difficulty with PVs was related to the way of teaching PVs in the classrooms. When teaching PVs, teachers tend to regard PVs as fixed phrases and require students to memorise the meanings of PVs by rote (Wei, 2021). The findings from the course book also support Wei's viewpoint. In the course book, students were provided with the specific meaning of a PV, without the non-contiguous use of separable PVs (see Appendix 4). As a result, language teachers are advised to rethink the way of teaching PVs to Chinese EFL learners. It is not enough to focus solely on the meanings of PVs; learners need to understand their actual use beyond meanings. Also, material designers should make it clearer that separable PVs have both contiguous and non-contiguous forms and encourage learners to use the non-contiguous form, especially when pronouns are used to be the object of separable PVs (e.g. *TAKE it out*, *TURN them off*).

Secondly, the findings of *LOOK up* in the Chinese learner corpus suggest the need for a more comprehensive list of English PVs. Due to the polysemous nature of PVs, the ready-made lists, such as the PHaVE list created by Garnier and Schmitt (2015), give

priority to the most frequent meanings of PVs, excluding less frequent meanings. While high frequency is an indicator of importance and these lists have largely relieved learners' pressure of learning various meanings of PVs, the ignorance of less frequent meanings in their work tends to be a potential weakness. For example, there was no corpus evidence showing that Chinese EFL learners were able to use the metaphorical meaning of *LOOK up* (*LOOK up*$_2$). Underestimating the importance of less frequent meanings may cause confusion, especially when learners encounter them in reading and listening. Moreover, the findings also show that Chinese EFL learners make mistakes when they use the non-compositional and idiomatic meaning of *LOOK up* (i.e. *LOOK up*$_4$). Based on the findings, language learners need a more comprehensive list of PVs that takes the metaphorical and idiomatic meanings of PVs into consideration. While the inclusion of more meanings of PVs will inevitably increase the burden on students, teachers could help by giving explicit instruction on the grammatical patterns associated with these PVs. I will elaborate on this point in the next section (Section 9.2.2).

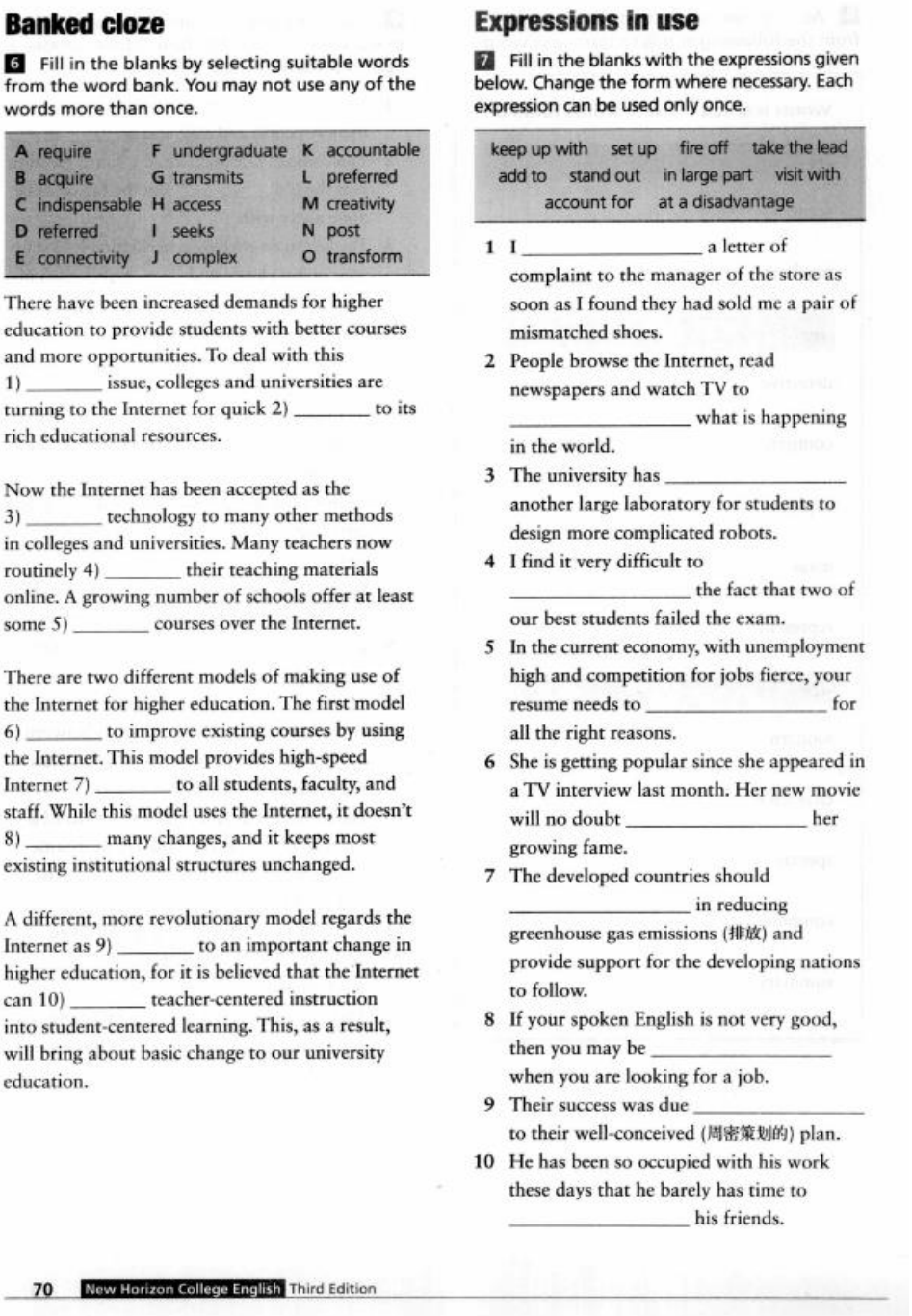
The results of three-word PVs have two important implications. One of the implications is that Chinese EFL learners tend to rely on their previous knowledge of PVs and lack exposure to idiomatic PVs. The findings from the current study (Section 9.1.2.2) and Gilquin's (2015) research suggest that learners stick to using the ones they are familiar with (e.g. *LOOK down upon* and *GET along with*) but ignore the ones with a higher degree of idiomaticity (e.g. *LOOK down on* and *GET on with*). This is probably due to the lack of exposure to idiomatic PVs. Learners may have repeated exposure to familiar PVs, using them more frequently and confidently. Teachers should be aware of learners' habitual choice and provide them with ample exposure to unfamiliar PVs, reducing their concern about idiomatic PVs. Another implication is associated with learners'

insensitivity to stylistic aspects of PVs. Chinese EFL learners seem to be confused about the formal and informal styles in written discourse. Learners use more three-word PVs in writing, yet they also prefer one-word verbs over PVs at the same time. This suggests that learners are not sensible to the stylistic differences between written and spoken discourse. In the classrooms, teachers can conduct some comparative activities and give explicit instruction on the stylistic differences between PVs and one-word verbs. Teachers can provide students with authentic examples extracted from both written and spoken texts, such as newspaper articles and interviews, or academic essays and conversation transcripts, and ask students to discuss the differences in word choice between the two texts.

9.2.2 Teaching PVs through phraseology

The phraseological knowledge of lexical items helps EFL learners produce language naturally and fluently (Pawley and Syder, 1983; Nattinger and DeCarrico, 1992). It would be beneficial to students to study the phraseology of PVs as well as the various meanings. PVs usually have multiple meanings and this polysemous feature often prohibit learners from using this language structure both in spoken and written discourse (Sinclair, 1996; Siyanova and Schmitt, 2007). As discussed above, it is inevitable that students will face greater challenges if they are taught the metaphorical and idiomatic meanings of PVs. Based on the findings of PVs in the BoE corpus, I argue that teaching PVs through collocation and grammatical patterns would reduce the burden of learning. Lexico-grammatical patterns help to differentiate the various meanings of polysemous items (Hunston and Francis, 2000). The findings also show that different meanings of *LOOK up* co-occur with different collocates and different grammatical patterns (see Section 6.1). Explicit instruction on the lexical and

grammatical patterns associated with PVs would make the process of learning PVs more effective and memorable. However, the investigation of PVs in the course book suggests that teachers and material designers have not yet realised the importance of phraseology.



Banked cloze

6 Fill in the blanks by selecting suitable words from the word bank. You may not use any of the words more than once.

A require	F undergraduate	K accountable
B acquire	G transmits	L preferred
C indispensable	H access	M creativity
D referred	I seeks	N post
E connectivity	J complex	O transform

There have been increased demands for higher education to provide students with better courses and more opportunities. To deal with this 1) _____ issue, colleges and universities are turning to the Internet for quick 2) _____ to its rich educational resources.

Now the Internet has been accepted as the 3) _____ technology to many other methods in colleges and universities. Many teachers now routinely 4) _____ their teaching materials online. A growing number of schools offer at least some 5) _____ courses over the Internet.

There are two different models of making use of the Internet for higher education. The first model 6) _____ to improve existing courses by using the Internet. This model provides high-speed Internet 7) _____ to all students, faculty, and staff. While this model uses the Internet, it doesn't 8) _____ many changes, and it keeps most existing institutional structures unchanged.

A different, more revolutionary model regards the Internet as 9) _____ to an important change in higher education, for it is believed that the Internet can 10) _____ teacher-centered instruction into student-centered learning. This, as a result, will bring about basic change to our university education.

Expressions in use

7 Fill in the blanks with the expressions given below. Change the form where necessary. Each expression can be used only once.

keep up with set up fire off take the lead
add to stand out in large part visit with
account for at a disadvantage

1 I _____ a letter of complaint to the manager of the store as soon as I found they had sold me a pair of mismatched shoes.
2 People browse the Internet, read newspapers and watch TV to _____ what is happening in the world.
3 The university has _____ another large laboratory for students to design more complicated robots.
4 I find it very difficult to _____ the fact that two of our best students failed the exam.
5 In the current economy, with unemployment high and competition for jobs fierce, your resume needs to _____ for all the right reasons.
6 She is getting popular since she appeared in a TV interview last month. Her new movie will no doubt _____ her growing fame.
7 The developed countries should _____ in reducing greenhouse gas emissions (排放) and provide support for the developing nations to follow.
8 If your spoken English is not very good, then you may be _____ when you are looking for a job.
9 Their success was due _____ to their well-conceived (周密策划的) plan.
10 He has been so occupied with his work these days that he barely has time to _____ his friends.

Figure 9.1 PVs exercise in *New Horizon College English – Reading and Writing: Student Book* (Zheng, 2015)

In the course book, PVs are regarded as fixed phrases, and the instruction focuses solely

on the meanings (see Appendix 4 for the instruction of PVs in the course book). Also, as Figure 9.1 above shows, exercises designed for PVs only assess learners' lexical knowledge of PVs, ignoring the phraseological use associated with them.

Because of the distinction between learners' L1 and L2, Chinese learners often make mistakes when using PVs, such as the ignorance of prepositions (see Section 8.2.1). Mueller (2011) advised that learning through collocation could effectively reduce errors with prepositions. With repeated exposure to the collocation and grammatical patterns that frequently occur in the co-text of PVs, learners are likely to use PVs more fluently and proficiently. As a result, teachers and course book designers are advised to pay more attention to the phraseology associated with PVs, especially collocation and grammatical patterns. Moreover, paying attention to the phraseology of PVs in the classrooms is also helpful to address learners' avoidance behaviour of using PVs. The findings from the current study and Hulstijn and Marchena's (1989) study show that learners tend to avoid PVs and regard one-word verbs as reliable replacements. However, corpus evidence extracted from the BoE corpus shows that replacing PVs with their one-word equivalents is unreliable because they often occur in different patterns. Chinese learners pay more attention to the semantic equivalence but ignore the grammatical differences between PVs and one-word verbs. Ind the classrooms, emphasizing the importance of patterning would be beneficial, as it helps students understand the distinction between PVs and one-word equivalents.

9.3 Contributions and limitations of the study

One of the contributions of the current study is to use semantic criteria to identify and

classify PVs in the corpus. The traditional method identifies PVs via syntactic tests (e.g. Bolinger, 1971). This method has potential weaknesses. For example, it fails to identify inseparable PVs such as *COME across* and *LOOK into* because they do not allow the movement of the particle. Unlike the traditional method, the current study used two semantic criteria to identify PVs (see Section 5.1). Using semantic criteria has some benefits. Firstly, it does not predefine the grammatical function of the non-verbal component but focuses on the association between the meaning of PVs and their components in the context. Secondly, the semantic criteria not only identify two-word PVs but also PVs with three components (i.e. phrasal-prepositional verbs) (see Section 5.3). Therefore, compared with the traditional method, using semantic criteria can identify a wider range of PVs.

Another contribution of the current study is to differentiate PVs and their one-word equivalents based on grammatical patterns. Previous studies have pointed out the stylistic differences between PVs and one-word verbs (e.g. Siyanova and Schmitt, 2007). In addition to the stylistic differences, this study provides corpus evidence to show the grammatical differences between them. It is unreliable to use one-word verbs to replace PVs because they are used in different grammatical patterns. However, Chinese EFL learners are not sensitive to these differences. Knowing the grammatical patterns associated with PVs and one-word verbs is beneficial to students because it helps students use both PVs and one-word verbs more naturally, reducing the error caused by the difference between learners' L1 and L2. For example, knowing the grammatical pattern associated with *LOOK up*$_4$ (e.g. **V n *in* n**) may help learners use this PV more proficiently.

While the current study provides insights into the use of PVs based on corpus evidence, it is important to acknowledge some limitations. First of all, this study focuses on a limited number of PVs. When I investigated how experienced users of English use PVs and identified the lexico-grammatical patterns associated with PVs in the BoE corpus, I paid particular attention to PVs that contained the word *LOOK.* Although *LOOK* is one of the high-frequency verbs in English and it is frequently combined with particles such as *up, down, into*, etc. to form PVs, the number of PVs remains to be a very small range. Findings on the use of PVs such as *LOOK up* provide supporting evidence to show that the patterns associated with *LOOK up* help to distinguish the different meanings of this combination. However, additional evidence from other PVs is expected in order to make a stronger argument that lexico-grammatical patterns are significant indicators of the different meanings of polysemous PVs. Moreover, the comparison on the use of PVs in the reference corpus and the Chinese learner corpus was based on a list of 100 frequent PVs proposed by Gardner and Davis (2007). Gardner and Davis (2007) observed that these PVs are mainly constructed by 8 particles, including *up*, *down*, *out*, *back*, *on*, *in*, *off*, *over*. Although the high frequency often signifies the importance of these PVs, many frequently occurring PVs are absent from this list, especially those that are constructed by other particles such as *away*, *through*, *across*, etc. In other words, the current study does not focus on the use of PVs such as *take away, go through, come across*, etc. in the two corpora because they are not included in Gardner and Davis's list. Due to the limited number of PVs included, the current study may not fully capture the overall use of PVs in the reference corpus and the Chinese learner corpus. This constraint could affect the generalizability of the findings to other PVs that were not taken into account.

When I investigated the frequency of 100 PVs, I categorised these PVs into 8 groups by particle and compared the frequency of each group between the learner corpus and the reference corpus. The grouping approach provided a systematic way to organize the large number of PVs, making them easier to analyse and compare. However, this grouping approach has some limitations and loses details such as the specific frequencies of PVs in the corpora. The analysis only shows the overall frequency of each group and assumes that each PV in a given group may have same frequency in the corpus. In other words, grouping PVs by particle ignores the specific frequency of one PV in the group. In the learner corpus, if a small number of PVs in the 'up' group are frequent, while others are very infrequent, it may not be reasonable to argue that PVs in the 'up' group are more frequent than PVs in the 'down' group. As a result, it is important to note that the current study provides a general overview of PV frequency across groups, without offering detailed frequencies for individual PVs within each group. The answers to the second research question do not suggest that all PVs in the 'up' group are more frequent than PVs in other groups in the Chinese learner corpus.

The corpora used in this study are not perfect with regard to the size and the inclusion of text files. Although I have made efforts to keep the two corpora as comparable as possible, the reference corpus is still larger than the learner corpus. This difference in size may result in potential bias, limiting the accuracy of comparisons between the two corpora. Moreover, the current study was based on two ready-made Chinese learner corpora (i.e. the WECCL and the TECCL corpus) and I did not collect learners' written texts on my own. Although both corpora aimed to reflect the English proficiency (especially the writing proficiency) of Chinese students who speak English as a foreign language (Wen, Liang and Yan, 2008; Xu, 2015), the texts may not fully reflect learners'

authentic language use in written discourse. This is because a large number of texts were written under controlled environments, such as exams and classroom essays. Since the texts were based on specific writing tasks, the language use may lack diversity. For example, *LOOK up*$_4$ is frequently used in the Chinese learner corpus because this PV is closely associated with one of the writing tasks (see topic 22 in Appendix 3). As a result, due to the limited range of writing tasks, the Chinese learner corpus only captures learners' use of some PVs but tend to ignore the use of other PVs. Furthermore, the study focuses exclusively on written texts, rather than spoken language. This narrow focus means that the findings may not be applicable to spoken discourse, where phrasal verb usage may differ significantly.

When I analysed the quantitative data collected from the corpora, I used the chi-square test to determine whether the different use of PVs in the learner corpus and the reference corpus was statistically significant. It may not be very surprising that learners' use of PVs was significantly different from L1 users' because the result was based on two large corpora. With large datasets, a small difference between observed and expected frequencies could be statistically significant (Gries, 2024). Also, the quantitative data did not account for potential differences between individual learners and individual texts. For example, while the quantitative analysis showed that *LOOK up* was frequently used by Chinese learners in their written texts, it did not show whether the high frequency of this PV was associated with an individual text or a specific essay topic. As a result, the current study showed the overall frequency of PVs and the use of specific PVs (e.g. *LOOK up, LOOK for* and *LOOK forward to*) in learners' written text, but it also has the limitation that it did not focus on individual differences.

Finally, the study focused on 10 PVs and their corresponding one-word equivalents. The 10 one-word verbs were chosen because they were closely connected with the most frequent meaning of the 10 PVs. Given the polysemous nature of PVs, it is clear that a PV does not have one potential one-word alternative. For example, *GO on* has multiple meanings and the one-word alternative *continue* is connected with only one meaning of the PV. It is also probable to claim that *GO on* can be substituted by *happen* in the following example:

This gives everyone the chance to find out what's ***going on***. (BNC: HCL)

Moreover, although learners strategically avoid using PVs by adopting one-word verbs, it is hard to anticipate their choice of replacement. Therefore, the finding that Chinese learners prefer to use one-word verbs over PVs is likely to be biased. The current study focused primarily on one-word alternatives, but it should be aware that there are other techniques to replace PVs, such as multi-word combinations (cf. Cornell, 1985).

9.4 Perspectives for further studies

Despite these limitations, the study sheds some light on the lexico-grammatical patterns associated with specific PVs. The findings in the current study show that it is possible to extend the investigation into patterning from the level of individual words to the level of semi-fixed phrases, such as PVs. English PVs are highly polysemous, but the current study has shown that the patterns in the co-text help to differentiate and understand the multiple meanings of polysemous PVs. Further studies could follow this approach and investigate the collocation and grammatical patterns associated with more frequent PVs. PVs usually have idiomatic meanings which cannot be predicted by the meanings of the components. However, this study has shown that the meaning of PVs has semantic

association with the frequent collocates in the co-text. For instance, corpus evidence shows that *LOOK up*$_1$ frequently co-occurs with *sky* and *stars*, whereas *LOOK up*$_4$ frequently co-occurs with *word* and *dictionary*. Knowing the frequent collocates helps to understand the distinctive meanings of PVs. Moreover, this study also shows the challenge of replacing PVs with one-word verbs. Although the meaning of the one-word verb is semantically equivalent with the PV, they are often used in different grammatical patterns (e.g. *LOOK for* and *search*). Corpus evidence suggests that using one-word verbs to replace PVs cannot be a 'play-if-safe' strategy if learners ignore the distinctive grammatical usages between PVs and one-word verbs. Future research could also focus on the grammatical patterns associated with PVs and their corresponding one-word verbs, providing more empirical evidence that PVs cannot be directly substituted by one-word equivalents.

Understanding PVs from the perspective of phraseology would also contribute to the learning of PVs. Traditionally, Chinese learners regard PVs as fixed phrases, memorizing the idiomatic meanings of PVs mechanically (Wei, 2021). It is argued in the current study that learning PVs through collocation and grammatical patterns would be beneficial to Chinese EFL learners, enhancing their phraseological knowledge of PVs. Due to the lack of PVs in their L1, learners struggle with using PVs naturally and fluently. If the phraseological patterns associated with frequent PVs are made available to them, they would be motivated to use PVs more naturally and proficiently. The phraseological knowledge of PVs also develops learners' mental lexicon of PVs. Future studies could go further and explore the way of learning and teaching PVs through collocation and grammatical patterns. Alternatively, researchers could also make a contrastive analysis and examine whether learning PVs via a lexico-grammatical

approach is more effective than the traditional approach.

The current study compares the different use of frequent PVs in the L1 users' writing and Chinese EFL learners' writing. The results have shown that Chinese EFL learners have difficulties using PVs proficiently in written texts due to the lack of phraseological knowledge of PVs. In addition to this, it is also valuable to know whether learners' use of PVs is influenced by other factors. For example, researchers could explore whether advanced learners have more phraseological knowledge of PVs than intermediate learners. It would be also interesting to know whether the writing task could be a factor that influences learners' use of PVs. Furthermore, this study focuses on the use of PVs in non-academic written texts. Considering this limitation of the current study, future studies may address this constraint by including learners' spoken corpora and incorporating spoken language analysis. In addition, it remains to be unknown whether Chinese learners are able to use PVs in academic writing and whether they have difficulties using PVs in academic context.

Finally, as shown in the current study, corpus data provides important information about how L1 users use specific PVs, which can be used to develop teaching materials. For example, when designing the teaching materials of three-word PVs, teachers and practitioners could incorporate with the frequency data, giving priority to *get on with* rather than *get along with* because the former is more frequent in the reference corpus. Course book designers can make full use of corpus evidence and design the materials based on the authentic data. As shown in the current study, concordance lines are very useful resources to observe the phraseological patterns associated with PVs and other lexical items. Teachers could use concordance lines in classroom activities, guiding

students to independently identify the phraseological patterns associated with PVs from the authentic data. In order to enhance the effectiveness of learning PVs, further studies could explore how to integrate corpus data with teaching materials.

References

Alangari, M., Jaworska, S., and Laws, J. (2020) Who's afraid of phrasal verbs? The use of phrasal verbs in expert academic writing in the discipline of linguistics. *Journal of English for Academic Purposes*, 43, pp. 1 – 13. https://doi.org/10.1016/j.jeap.2019.100814

Altenberg, B. and Granger, S. (2001) The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22 (2), pp. 173 – 195. https://doi.org/10.1093/applin/22.2.173

Anthony, L. (2023) AntConc (Version 4.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Barnbrook, G. (2009) 'Sinclair on collocation', in R. Moon (ed.) *Words, grammar, text: revisiting the work of John Sinclair*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 23 – 38.

Barnbrook, G., Mason, O., and Krishnamurthy, R. (2013). *Collocation: applications and implications.* London: Palgrave Macmillan.

Becker, T. P. (2014) Avoidance of English phrasal verbs: investigating the effect of proficiency, learning context, task type, and verb type. *Asian Journal of English Language Teaching*, 24, pp. 1 – 33. https://muse.jhu.edu/article/562647.

Biber, D., Conrad, S., and Cortes, V. (2004) If you look at…: lexical bundles in university teaching and textbooks. *Applied linguistics*, 25 (3), pp. 371 – 405. https://doi.org/10.1093/applin/25.3.371

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

Bolinger, D. (1971) *The phrasal verb in English*. Cambridge (Mass.)/London: Harvard University Press.

Brezina, V. (2018) *Statistics in corpus linguistics: a practical guide*. Cambridge: Cambridge University Press.

Burnard, L. (2007) *Reference guide for the British National Corpus (XML Edition),* Available at: http://www.natcorp.ox.ac.uk/docs/URG/

Carter, R. (2012) *Vocabulary: applied linguistic perspectives*. New York/London: Routledge.

Celce-Murcia, M. and Larsen-Freeman, D. (1999) *The grammar book: an ESL/EFL teacher's course.* Boston, MA: Heinle & Heinle.

Che, S., Chen, J., and Jiang, Z. (2024) National consciousness in college English course books: study of New Horizon College English Reading and Writing. *Contemporary Foreign Languages Studies*, 24 (4), pp. 73 – 82.

Chen, J. (2007) On how to solve the problem of the avoidance of phrasal verbs in the Chinese context. *International Education Journal*, 8 (2), pp. 348 – 353.

Chen, M. (2013) Overuse or underuse: a corpus-based study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, 18 (3), pp.418 – 442. https://doi.org/10.1075/ijcl.18.3.07che

Clear, J. (1993) 'From Firth principles: computational tools for the study of collocation', in M. Baker, G. Francis, and E. Tognini-Bonelli (eds.) *Text and technology: in honour of John Sinclair*. Philadelphia: John Benjamins Publishing Company, pp. 271 – 292.

Cornell, A. (1985) Realistic goals in teaching and learning phrasal verbs.

International Review of Applied Linguistics in Language Teaching, 23 (4), pp. 269 – 280. https://doi.org/10.1515/iral.1985.23.1-4.269

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied linguistics*, 2 (3), pp. 223 – 235. https://doi.org/10.1093/applin/II.3.223

Cowie, A. P. (1988) 'Stable and creative aspects of vocabulary use', In R. Carter and M. J. McCarthy (eds.) *Vocabulary and language teaching*. London: Longman, pp. 126 – 137.

Cowie, A. P. (1998) *Phraseology: theory, analysis, and applications.* Oxford: Clarendon Press.

Cowie, A. P. and Mackin, R. (1993) *Oxford dictionary of phrasal verbs*. Oxford: Oxford University Press.

Crystal, D. (2004) *Making sense of grammar.* Harlow: Pearson Longman.

Darwin, C. M. and Gray, L. S. (1999) Going after the phrasal verbs: an alternative approach to classification. *TESOL Quarterly*, 33 (1), pp. 65 – 83. https://doi.org/10.2307/3588191

Downing, A. and Locke, P. (2002) *A university course in English grammar*. New York and London: Routledge.

Durrant, P. (2009) Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28 (3), pp. 157 – 169. https://doi.org/10.1016/j.esp.2009.02.002

Durrant, P. (2014) Corpus frequency and second language learners' knowledge of collocations: a meta-analysis. *International Journal of Corpus Linguistics*, 19 (4), pp. 443 – 477. https://doi.org/10.1075/ijcl.19.4.01dur

Durrant, P. and Mathews-Aydinli, J. (2011) A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, 30 (1), pp.58 – 72. https://doi.org/10.1016/j.esp.2010.05.002

Durrant, P. and Schmitt, N. (2009) To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47 (2), pp. 157 – 177. https://doi.org/10.1515/iral.2009.007

Durrant, P., Siyanaova-Chanturia, A., Kremmel, B., and Sonbul, S. (2022) *Research methods in vocabulary studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Evert, S. (2009) 'Corpora and collocations', in A. Lüdeling and M. Kytö (eds.) *Corpus linguistics: an international handbook*. Berlin: Mouton de Gruyter, pp. 1212 – 1248.

Francis, G. (1993) 'A corpus-driven approach to grammar: principles, methods and examples', in M. Baker, G. Francis, and E. Tognini-Bonelli (eds.) *Text and technology: in honour of John Sinclair*. Philadelphia: John Benjamins Publishing Company, pp. 137 – 156.

Francis, G., Hunston, S., and Manning, E. (1996) *Collins COBUILD grammar patterns 1: verbs*. London: HarperCollins.

Francis, G., Hunston, S., and Manning, E. (1996) *Collins COBUILD grammar patterns 2: nouns and adjectives*. London: HarperCollins.

Fraser, B. (1976) *The verb-particle combination in English*. New York, San Francisco and London: Academic Press.

Gablasova, D., Brezina, V., and McEnery, T. (2017) Exploring learner language through corpora: Comparing and interpreting corpus frequency information.

Language Learning, 67 (S1), pp. 130 – 154. https://doi.org/10.1111/lang.12226

Gardner, D. and Davies, M. (2007) Pointing out frequent phrasal verbs: a corpus-based analysis. *TESOL Quarterly*, 41 (2), pp. 339 – 359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x

Gardner, D. and Davis, M. (2018) Sorting them all out: exploring the separable phrasal verbs of English. *System*, 76, pp. 197 – 209. https://doi.org/10.1016/j.system.2018.06.009

Garnier, M. and Schmitt, N. (2015) The PHaVE list: a pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19 (6), pp. 645 – 666. https://doi.org/10.1177/1362168814559798

Garnier, M. and Schmitt, N. (2016) Picking up polysemous phrasal verbs: how many do learners know and what facilitates this knowledge? *System*, 59, pp. 29 – 44. http://dx.doi.org/10.1016/j.system.2016.04.004

Gilquin, G. (2015) The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistics Theory*, 11 (1), pp. 51 – 88. https://doi.org/10.1515/cllt-2014-0005

Gilquin, G. (2023) Second and foreign language learners: the effect of language exposure on the use of English phrasal verbs. *International Journal of Bilingulism*, 1 (8), pp. 1 – 18. https://doi.org/10.1177/13670069231169123

Granger, S. (1998) 'Prefabricated patterns in EFL writing', in A. P. Cowie (ed.) *Phraseology: theory, analysis, and applications*. Oxford: Clarendon Press, pp. 145 – 160.

Granger, S. (2021) 'Phraseology, corpora, and L2 research', in S. Granger (ed.)

Perspectives on the L2 phrasicon: the view from learner corpora. Bristol/Blue Ridge Summit: Multilingual Matters, pp. 3 – 24.

Granger, S. and Bestgen, Y. (2014) The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *Language Teaching,* 52 (3), pp. 229 – 252. https://doi.org/10.1515/iral-2014-0011

Granger, S. and Meunier, F. (2008) 'Introduction: the many faces of phraseology' in S. Granger. and F. Meunier. (eds.) *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. XIX – XXVIII.

Granger, S. and Paquot, M. (2008) 'Disentangling the phraseological web', in S. Granger. and F. Meunier. (eds.) *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 27 – 49.

Gries, S. Th. (1999) Particle movement: a cognitive and functional approach. *Cognitive Linguistics*, 10 (2), pp. 105 – 145. https://doi.org/10.1515/cogl.1999.005

Gries, S. Th. (2009) *Quantitative corpus linguistics with R: a practical introduction*. New York and London: Routledge.

Gries, S. Th. (2010) Useful statistics for corpus linguistics. *A mosaic of corpus linguistics: Selected approaches*, 66, pp. 269 – 291.

Gries, S. Th. (2024) *Frequency, dispersion, association, and keyness: Revising and tupelizing corpus-linguistic measures*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gries, S. Th. and Durrant, P. (2021) 'Analysing co-occurrence data', in M. Paquot and S. Th. Gries (eds.) *A practical handbook of corpus linguistics*. Cham: Springer International Publishing, pp. 141 – 159.

Gries, S. Th. and Stefanowitsch, A. (2004) Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9 (1), pp. 97 – 129. https://doi.org/10.1075/ijcl.9.1.06gri

Groom, N. (2005) Pattern and meaning across genres and disciplines: an exploratory study. *Journal of English for Academic Purposes*, 4, pp. 257 – 277. https://doi.org/10.1016/j.jeap.2005.03.002

Guo, L. and Chen, Y. (2007) A comparative study of CET-4 and TEM-4 in China, *Asian Englishes*, 10 (1), pp. 90 – 102. https://doi.org/10.1080/13488678.2007.10801201

Hardie, A. (2012) CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17 (3), pp. 380 – 409. https://doi.org/10.1075/ijcl.17.3.04har

Hoey, M. (2005) *Lexical priming: a new theory of words and language*. New York and London: Routledge.

Howarth, P. A. (1996) *Phraseology in English academic writing*. Tübingen, Germany: Walter de Gruyter.

Hunston, S. (2001) 'Colligation, lexis, pattern, and text', in M. Scott and G. Thompson (eds.) *Patterns of text: in honour of Michael Hoey*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 13 – 33.

Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2007) 'Phraseology and system a contribution to the debate', in G. Thompson and S. Hunston (eds.) *System and corpus: exploring connections.* London: Equinox, pp. 55 – 80.

Hunston, S. (2008) Starting with small words: patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13 (3), pp. 271 – 295. https://doi.org/10.1075/ijcl.13.3.03hun

Hunston, S. (2010) 'How a corpus be used to explore patterns?' in A. O'Keeffe and M. McCarthy (eds.) *The Routledge handbook of corpus linguistics*. New York and London: Routledge, pp. 152 – 166.

Hunston, S. (2014) 'Pattern grammar in context', in T. Herbst, H. Schmid, and S. Faulhaber (eds.) *Constructions, collocations, patterns.* Berlin: De Gruyter Mouton, pp. 99 – 120.

Hunston, S. and Francis, G. (2000) *Pattern grammar: a corpus-driven approach to lexical grammar of English.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hulstijn, J. H. and Marchena, E. (1989) Avoidance: grammatical or semantic cause? *Studies in Second Language Acquisition*, 11 (3), pp. 241 – 255. https://doi.org/10.1017/S0272263100008123

Hyland, K. (2008) As can be seen: lexical bundles and disciplinary variation. *English for specific purposes,* 27 (1), pp. 4 – 21. https://doi.org/10.1016/j.esp.2007.06.001

Kövecses, Z. (2010) *Metaphor: a practical introduction*. 2nd edition. Oxford/New York: Oxford University Press.

Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. Chicago/London: University of Chicago press.

Laufer, B. and Waldman, T. (2011) Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language Learning*, 61 (2), pp. 647 – 672. https://doi.org/10.1111/j.1467-9922.2010.00621.x.

Li, F., Ren, J., and Zhao, H. (2016) Grammatical mistakes in college English writing: problem analysis, reasons and solutions. *International Journal of Applied Linguistics and Translation*, 2 (3), pp. 20 – 28. Doi: 10.11648/j.ijalt.20160203.11.

Liao, Y. and Fukuya, Y. J. (2004) Avoidance of phrasal verbs: the case of Chinese learners of English. *Language Learning*. 54 (2), pp. 193 – 226. https://doi.org/10.1111/j.1467-9922.2004.00254.x

Lindley, J. (2020) Discourse functions of *always* progressives: beyond complaining. *Corpus Linguistics and Linguistics Theory.* 16 (2), pp. 333 – 361. https://doi.org/10.1515/cllt-2016-0028

Lindstromberg, S. (2022) The compositionality of English phrasal verbs in terms of imageability. *Lingua*, 275, pp. 1 – 25. https://doi.org/10.1016/j.lingua.2022.103373

Liu, D. (2010) Going beyond patterns: involving cognitive analysis in the learning of collocations. *TESOL Quarterly,* 44 (1), pp. 4 – 30. https://doi.org/10.5054/tq.2010.214046

Liu, D. (2011) The most frequently used English phrasal verbs in American and British English: a multicorpus examination. *TESOL Quarterly*, 45 (4), pp. 661 – 668. https://doi.org/10.5054/tq.2011.247707

Liu, D. and Myers. D. (2020) The most-common phrasal verbs with their key meanings for spoken and academic written English: a corpus analysis. *Language Teaching Research*, 24 (3), pp. 403 – 424. https://doi.org/10.1177/1362168818798384

Louw, B. (1993) 'Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosody', in M. Baker, G. Francis, and E. Tognini-

Bonelli (eds.) *Text and technology: in honour of John Sinclair.* Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 157 – 176.

Louw, B. (2000) 'Contextual prosodic theory: bringing semantic prosodies to life', in C. Heffer. and H. Sauntson. (eds.) *Words in context: a tribute to John Sinclair on his retirement.* Birmingham: University of Birmingham, pp. 48 – 94.

Lu, Z. and Sun, J. (2017) Presenting English polysemous phrasal verbs with two metaphor-based cognitive methods to Chinese EFL learners. *System*, 69, pp. 153 – 161. https://doi.org/10.1016/j.system.2017.07.016

Jin, Y. and Fan, J. (2011) Test for English majors (TEM) in China. *Language Testing*, 28 (4), pp. 589 – 596. https://doi.org/10.1177/0265532211414852

Mahlberg, M. (2005) *English general nouns: a corpus theoretical approach*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Martinez, R. and Schmitt, N. (2012) A phrasal expressions list. *Applied Linguistics*, 33 (3), pp. 299 – 320. https://doi.org/10.1093/applin/ams010

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

McEnery, T. and Wilson, A. (1996) *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Milton, J. and Hopwood, O. (2022) *Vocabulary in the foreign language curriculum: principles for effective instruction*. New York: Routledge.

Mindt, D. (1997) 'Complementary distribution, gradience and overlap in corpora and in ELT: Analysing and teaching the progressive', in U. Fries, V. Muller, and P. Schneider (eds.) *From Ælfric to the New York Times: studies in English corpus*

linguistics, Amsterdam: Rodopi, pp. 227 – 237.

Mindt, D. (2000) *An empirical grammar of the English verb system*. Berlin: Cornelsen Verlag.

Moon, R. (1997) 'Vocabulary connections: multi-word items in English', in N. Schmitt and M. McCarthy (eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press, pp. 40 – 63.

Moon, R. (1998) *Fixed Expressions and Idioms in English.* Oxford: Clarendon Press.

Mueller, C. M. (2011) English learners' knowledge of prepositions: collocational knowledge or knowledge based on meaning? *System*, 39, pp. 480 – 490. https://doi.org/10.1016/j.system.2011.10.012.

Murphy, M. L. (2003) *Semantic relations and lexicon: antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.

Na, S. and Yoon, H. (2016) Effects of in-class and out-of-class writing assignments on L2 writing strategy use and writing quality. *The Asia-Pacific Education Researcher*, 25, pp. 195 – 205. https://doi.org/10.1007/s40299-015-0250-5

Nation, I. S. P (2013) *Learning vocabulary in another language.* 2nd Edition. Cambridge: Cambridge University Press.

Nation, P. and Waring, R. (1997) 'Vocabulary size, text coverage and word lists', in N. Schmitt and M. McCarthy (eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press, pp. 6 – 19.

Nattinger, J. R. and DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching.* Oxford: Oxford University Press.

Nesi, H., Gardner, S., Chapelle, C. A., and Hunston, S. (2012) *Genre across the disciplines: student writing in higher education.* Cambridge: Cambridge

University Press.

Nesselhauf, N. (2004) *Collocations in a learner corpus.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

O'Keeffe, A. and McCarthy, M. (2010) *The Routledge handbook of corpus linguistics*. New York and London: Routledge.

Oakes, M. P. (1998) *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Pawley, A. and Syder, F. H. (1983) 'Two puzzles for linguistic theory: nativelike selection and nativelike fluency', in J. C. Richards and R. W. Schmidt (eds.) *Language and communication.* London: Longman, pp. 191 – 226.

Princeton University (2010) *About WordNet*. Princeton University. Accessible at http://wordnet.princeton.edu.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J., (1985) *A comprehensive grammar of the English language.* London/New York: Longman.

Riguel, E. (2014) 'Phrasal verbs, "the scourge of the learner"', in *9th Lancaster University Postgraduate Conference in Linguistics & Language Teaching*, Vol. 9, pp. 1 – 20.

Rodríguez-Puente, P. (2019) *The English phrasal verb, 1650-present: history, stylistic drifts, and lexicalisation*. Cambridge: Cambridge University Press.

Römer, U. (2005) *Progressives, patterns, pedagogy: a corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Römer, U. (2009) The inseparability of lexis and grammar: corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7, pp. 141 – 163.

https://doi.org/10.1075/arcl.7.06rom

Rundell, M. (2008) The corpus revolution revisited. *English Today*, 24 (1), pp. 23 – 27. https://doi.org/10.1017/S0266078408000060

Rudzka-Ostyn, B. (2003) *Word power: phrasal verbs and compounds*. Berlin: Mouton de Gruyter.

Sakaba, H. and Okada, T. (2021) Usage patterns and meanings of high-frequency English verbs: a multi-word expression approach to Japanese high school EFL textbook analysis. *International Journal of Applied Linguistics & English Literature*, 10 (4), pp. 116 – 130. https://doi.org/10.7575/aiac.ijalel.v.10n.4p.116

Schmitt, N. (2000) *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. and Carter, R. (2004) *Formulaic sequences: acquisition, processing and use*. Amsterdam/Philadelphia: John Benjamins Publishing Company

Schmitt, N. and Redwood, S. (2011) 'Learner knowledge of phrasal verbs: a corpus-informed study', in F. Meunier, S. De Cock, G, Gilquin, and M, Paquot (eds.) *A taste for corpora: In honour of Sylviane Granger*, pp. 173 – 208.

Schmitt, N. and Schmitt, D. (2014) A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47 (4), pp. 484 – 503. https://doi.org/10.1017/S0261444812000018

Side, R. (1990) Phrasal verbs: sorting them out. *ELT Journal*, 44 (2), pp.144 – 152. https://doi.org/10.1093/elt/44.2.144

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

Sinclair, J. (1996) The search for units of meaning. *Textus*, 9 (1), pp. 75 – 106.

Sinclair, J. (2004a) 'Lexis and grammar', in J. Sinclair and R. Carter (eds.) *Trust the text: language, corpus and discourse*. London: Routledge.

Sinclair, J. and Carter, R. (ed.). (2004b) *Trust the text: language, corpus and discourse*. London: Routledge.

Sinclair, J. (2008) 'Preface', in S. Granger and F. Meunier (eds.) *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. XV-XVIII.

Sinclair, J., Jones, S., and Daley, R. (2004) *English collocation studies: the OSTI report*. London: Continuum.

Siyanova, A. and Schmitt, N. (2007) Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching*, 45 (2), pp. 121 – 139. https://doi.org/10.1515/IRAL.2007.005

Siyanova-Chanturia, A. and Webb, S. (2016) 'Teaching vocabulary in the EFL context', in W. A. Renandya and H. P. Widodo (eds.) *English language teaching today: linking theory and practice.* Springer International Publishing Switzerland, pp. 227 – 239.

Stubbs, M. (1995) Collocation and semantic profiles: On the cause and trouble with quantitative methods. *Functions of language*, 2 (1), pp. 23 – 55. https://doi.org/10.1075/fol.2.1.03stu

Stubbs, M. (2001) *Words and phrases: corpus studies of lexical semantics.* Oxford: Blackwell Publishing.

Stubbs, M. (2009) 'Technology and phraseology: with notes on the history of corpus linguistics', in U. Römer and R. Schulze (eds.) *Exploring the lexis-grammar*

interface. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 15 – 32.

Su, H. (2019) Is a meaning-based grammar of English possible? Exploring English grammar from the perspectives of pattern grammar and local grammar. *English Today*, 35 (4), pp. 53 – 59. https://doi.org/10.1017/S0266078418000524

Sung, M. (2020) Underuse of English verb-particle constructions in an L2 learner corpus: focus on structural patterns and one-word preference. *Corpus Linguistics and Linguistics Theory.* 16 (1), pp. 189 – 214. https://doi.org/10.1515/cllt-2017-0002

Thim, S. (2012) *Phrasal verbs: the English verb-particle construction and its history*. Berlin & Boston: Walter de Gruyter.

Trebits, A. (2009) The most frequent phrasal verbs in English language EU documents – a corpus-based analysis and its implications. *System*, 37, pp. 470 – 481. https://doi.org/10.1016/j.system.2009.02.012

Waibel, B. (2007) *Phrasal verbs in learner English: a corpus-based study of German and Italian students* (Unpublished PhD thesis). Albert-Ludwigs-Universitat Freiburg, Freiburg.

Walker, C. (2011a) How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes*, 30 (2), pp. 101 – 112. https://doi.org/10.1016/j.esp.2010.12.003

Walker, C. (2011b) A corpus-based study of the linguistic features and processes which influence the way collocations are formed: some implications for the learning of collocations. *TESOL Quarterly*, 45 (2), pp. 294 – 312. https://doi.org/10.5054/tq.2011.247710

Wei, Y. (2021) Use of English phrasal verbs of Chinese students across proficiency levels: a corpus-based analysis. *International Journal of TESOL Studies*, 3 (4), pp. 25 – 41. http://doi.org/10.46451/ijts.2021.12.03

Wen, Q., Liang, M., and Yan, X. (2008) *Spoken and written English corpus of Chinese learners (version 2.0).* Beijing: Foreign Language Teaching and Research Press.

Wray, A. (2002) *Formulaic language and lexicon.* Cambridge: Cambridge University Press.

Wu, J. and Tissari, H. (2021) Intensifier-verb collocations in academic English by Chinese learners compared to native-speaker students. *Chinese Journal of Applied Linguistics*, 44 (4), pp. 470 – 487. https://doi.org/10.1515/CJAL-2021-0030

Wulff, S. and Baker, P. (2021) 'Analysing concordances' in M. Paquot and S. Th. Gries (eds.) *A practical handbook of corpus linguistics.* Springer International Publishing AG. pp. 161 – 179.

Wulff, S. and Gries, S. (2019) Particle placement in learner language. *Language Learning*, 69 (4), pp. 873 – 910. https://doi.org/10.1111/lang.12354

Xu, J. (2015) Ten-thousand English compositions of Chinese learners (version 1.1). The National Research Centre for Foreign Language Education, Beijing Foreign Studies University. Retrieved at: https://corpus.bfsu.edu.cn/info/1070/1449.htm (Accessed: 7 September 2024)

Xu, J. and Fan, Y (2017) The evolution of the college English curriculum in China (1985 – 2015): changes, trends and conflicts. *Language Policy*, 16, pp. 267 – 289. https://doi.org/10.1007/s10993-016-9407-1

Yasuda, S. (2010) Learning phrasal verbs through conceptual metaphors: a case of

Japanese EFL learners. *TESOL Quarterly,* 44 (2), pp. 250 – 273. https://doi.org/10.5054/tq.2010.219945

Zheng, S. (ed.) (2015) *New Horizon College English – Reading and Writing: Student Book*, 3rd edition. Beijing: Foreign Language Teaching and Research Press.

Zipp, L. and Bernaisch, T. (2012) 'Particle verbs across first and second language varieties of English', in M. Hundt and U. Gut (eds.) *Mapping unity and diversity world-wide: corpus-based studies in new Englishes*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.167 – 196.

Zou, B. and Reinders, H. (2017) 'Using corpora to investigate Chinese university EFL learners', in H. Reinders, D. Nunan, and B. Zou (eds.) *Innovation in language learning and teaching: the case of China*. London: Palgrave Macmillan, pp. 245 – 260.

Appendices

Appendix 1: Gardner and Davis's list of top 100 frequent PVs in the BNC corpus

1. **out** (19): break out, bring out, carry out, come out, find out, get out, give out, go out, hold out, look out, make out, move out, pick out, point out, put out, set out, take out, turn out, work out

2. **up** (15): break up, bring up, come up, get up, give up, go up, hold up, look up, make up, move up, pick up, put up, set up, sit up, take up, turn up

3. **back** (12): bring back, come back, get back, give back, go back, hold back, look back, move back, put back, sit back, take back, turn back

4. **down** (10): break down, bring down, come down, get down, go down, look down, put down, set down, sit down, take down

5. **in** (8): bring in, come in, get in, give in, go in, move in, put in, take in

6. **off** (8): break off, come off, get off, go off, put off, set off, take off, turn off

7. **on** (8): carry on, come on, get on, go on, hold on, move on, put on, take on

8. **over** (4): come over, go over, take over, turn over

9. **others** (14): bring about, come about, set about, come along, go along, look around, turn around, come round, go round, look round, turn round, come through, get through, go through

Appendix 2: Top 20 frequent three-word PVs in the reference corpus and the learner corpus

	PVs	Reference corpus		Learner corpus	
		Freq.	Freq. per million	Freq.	Freq. per million
1	LOOK forward to	138	32.29	295	102.34
2	TURN out to	88	20.59	43	14.92
3	MAKE up of	50	11.70	35	12.14
4	GET out of	42	9.83	49	17.00
5	DATE back to	34	7.96	16	5.55
6	bound up with	27	6.32	5	1.73
7	COME up with	26	6.08	87	30.68
8	LIVE up to	24	5.62	20	6.94
9	RUN out of	18	4.21	12	4.16
10	KEEP up with	18	4.21	115	39.90
11	CATCH up with	18	4.21	113	39.20
12	STAND up to	18	4.21	2	0.69
13	GO along with	17	3.98	10	3.47
14	PUT up with	15	3.51	18	6.24
15	GET on with	14	3.28	24	8.33
16	MAKE up for	11	2.57	16	5.55
17	GET away with	10	2.34	3	1.04
18	END up with	10	2.34	5	1.73
19	FACE up to	10	2.34	25	8.67
20	LOOK down (up)on	10	2.34	49	17.00
Total		598	139.93	942	326.81

Appendix 3: Topics of written texts in the WECCL (Wen, Liang, and Yan, 2008)

A. Argumentative essays

01. "Education is expensive, but the consequences of a failure to educate, especially in an increasingly globalized world, are even more expensive." Write an essay of approximately 300 words on this issue to state your own opinion.
02. Some people think that education is a life-long process, while others don't agree. Write an essay to state your own opinion.
03. Nowadays, people have paid more and more attention to degree certificates. For example, in many institutions, one's promotion is primarily decided by whether one has obtained a graduate degree or not. A growing number of critics say that if this tendency goes to the extreme, young people may be misled. A degree certificate can reflect only one's academic achievements but not all abilities essential for successful career. Write an essay of approximately 300 words on this issue to state your own opinion.
04. An African proverb says "If you educate a boy, you educate an individual; if you educate a girl, you educate a family and a nation." Do you agree with this proverb? Write an essay of approximately 300 words on this issue to state your own opinion.
05. Computer games are very popular among children. However, some people think that computer games have produced more negative effects than positive ones on children's physical, intellectual as well as psychological development. Therefore, they suggest that effective measures should be taken to prevent children from playing them. Write an essay to state your own opinion.
06. Does modern technology make life more convenient, or was life better when technology was simpler? Write an essay to state your own opinion.
07. Some people think that the animals should be treated as pets, while others think that animals are resources of food and clothing. What is your opinion?
08. In the western world, if a family member has got a cancer, his/her family members must tell him/her about it frankly. If not, it would be regarded as being illegal. But in the Chinese culture, a common practice is not to tell the patient the truth. Some people think that this traditional practice must be changed along with the development of modernization. Write an essay of approximately 300 words on this issue to state your own opinion.

09. It is right that college graduates earn higher salaries than the less well-educated in the community. But they should also pay the full cost of their study. Do you agree or disagree?
10. Some people think that famous people are treated unfairly by the media, and they should be given more privacy, while some others think that this is the price of their fame. Write an essay to state your own opinion.
11. Many people say that we have developed into a “throw-away society”, because we are filling up our environment with so many plastic bags and rubbish that we cannot fully dispose of. To want extent do you agree with this opinion and what measures can you recommend to reduce this problem.
12. Many people think that work nowadays is more stressful and less leisurely than in the past. What is your opinion?
13. Nowadays men are becoming more and more greedy and selfish. We should return to older, traditional values and show respect for family and local community. To what extent do you agree or disagree?
14. Nowadays, we are advised by environmentalists to use electronic cards instead of paper cards for holiday greetings. However, some people think that electronic cards do not have the same flavor of paper cards and do not display the same function, either. Write an essay to state your own opinion.
15. Some people say the government shouldn’t put money on building theaters and sports stadiums; they should spend more money on medical care and education. Do you agree or disagree? State the reasons for your view.
16. Nowadays senior high school students are totally tired of various kinds of examinations given by their teachers in preparing for the future college entrance examinations. It is generally agreed that this kind of examination system has destroyed students’ creative thinking abilities and hindered their all-round development. However, the views on how to remedy the situation are various. Some people suggest that this type of examination system should be abolished completely while others think the abolishment of the examination system will bring about more problems than solutions. For example, without a national entrance examination, we will have problems of privileges and discrimination. Write an essay of approximately 300 words on this issue to state your own opinion.
17. Some people think the university education is to prepare students for employment.

Others think it has other functions. Discuss and say what other functions you think it should have.

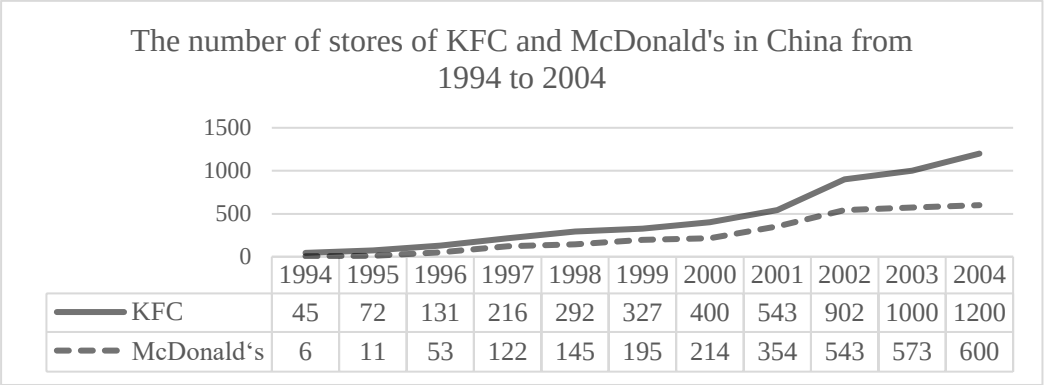
18. Some sport events such as the World Cup may help reduce the tension and bias between different countries and keep the peace of the world. What is your opinion?
19. The brain drain is a serious problem in developing countries. Some people think the reason for losing the most precious resources is the governments' poor policy. If the governments in the developing countries face up to the new reality, this problem can be alleviated. However, some people think that the brain drain is a universal phenomenon. No matter whatever measures the government takes, the problem of the brain drain cannot be solved. Write an essay of approximately 300 words on this issue to state your own opinion.
20. Traffic and housing problems in major cities would be solved by moving big companies, factories and their employees to the countryside. Do you agree or disagree?
21. Which skill of English is more important for Chinese learners? Some people think that we should give priority to reading in English, while others think speaking is more important. Write an essay to state your own opinion.
22. Nowadays, electronic dictionaries (E-dictionaries) have been increasingly popular among students. However, teachers think that the overuse of E-dictionaries might have more disadvantages than advantages for English learning. For example, like the use of calculator affecting the skill of calculating, reliance on E-dictionaries may lead to the deteriorating of our spelling ability. Write an essay of approximately 300 words on this issue to state your own opinion.
23. Some people think children should learn to compete, but others think that children should be taught to cooperate. Express some reasons of both views and give your own opinion.
24. Will modern technology, such as the internet ever replace the book or the written word as the main source of information? Write an essay to state your own opinion.
25. Young people are important resources to their country. But governments may ignore some problems faced by young people in running the country. By your experience, what do the government need to do for supporting or helping young people? Show these problems and give your ideas or suggestions to solve this issue.
26. Nowadays, more and more college students rent apartments and live outside campus.

Is it appropriate? State your opinion about this.

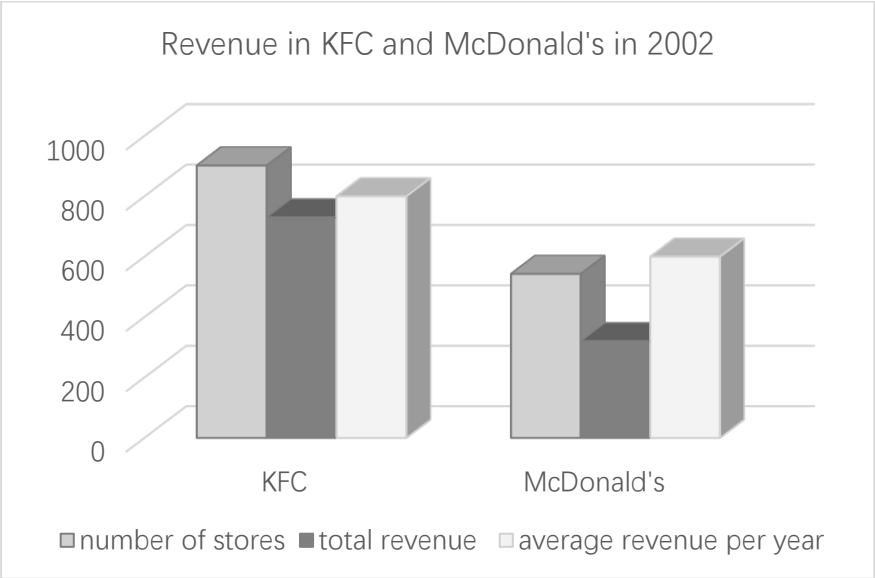
B. Expository essay

01. Writing Task (100 marks, 30 minutes)

You are going to give a presentation about the development of KFC and McDonald's over a ten-year period in China. Use the information in the following two graphs and write a report in English (150-180 words) for your presentation. Write your report on the separate answer sheet.

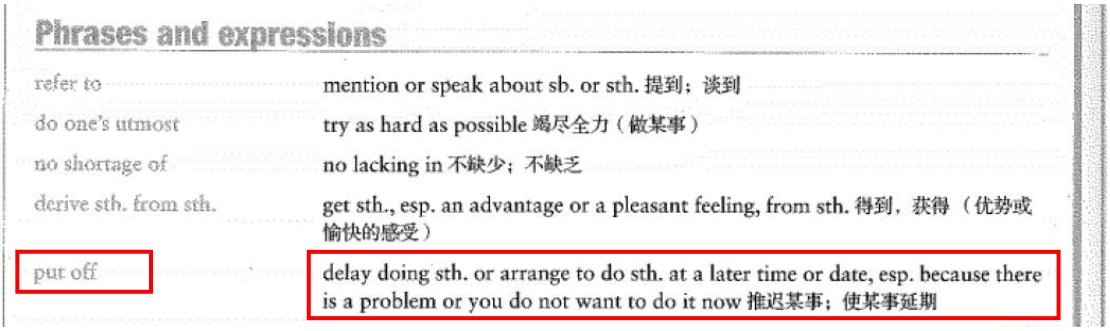


	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
KFC	45	72	131	216	292	327	400	543	902	1000	1200
McDonald's	6	11	53	122	145	195	214	354	543	573	600



Appendix 4: Instruction on two-word and three-word PVs in *New Horizon College English Reading and Writing: Student Book* (Zheng, 2015)

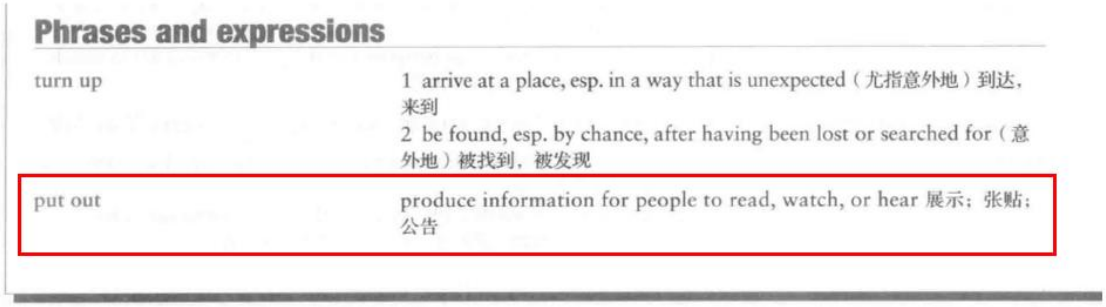
- ***PUT off***



Phrases and expressions

refer to	mention or speak about sb. or sth. 提到；谈到
do one's utmost	try as hard as possible 竭尽全力（做某事）
no shortage of	no lacking in 不缺少；不缺乏
derive sth. from sth.	get sth., esp. an advantage or a pleasant feeling, from sth. 得到，获得（优势或愉快的感受）
put off	delay doing sth. or arrange to do sth. at a later time or date, esp. because there is a problem or you do not want to do it now 推迟某事；使某事延期

- ***PUT out***



Phrases and expressions

turn up	1 arrive at a place, esp. in a way that is unexpected（尤指意外地）到达，来到 2 be found, esp. by chance, after having been lost or searched for（意外地）被找到，被发现
put out	produce information for people to read, watch, or hear 展示；张贴；公告

- ***CATCH up with***

Phrases and expressions

live in terror of sb. / sth.	be very afraid of sb. or sth. all the time 一直很害怕某人 / 某物
a string of	a number of similar things or events coming one after another 一连串，一系列（事件等）
turn off	(*infml.*) stop listening to or thinking about sb. or sth. 不再听；不再想
lean on	depend on sb. for support and encouragement, esp. at a difficult time 依靠；依赖
beat sb. to it	(*infml.*) do sth. before sb. else 抢在某人之前做
pull over	drive to the side of the road and stop your car, or make sb. else do this（使）（车辆或司机）停靠在路边
slip off / down	move smoothly, esp. off or from sth. 滑落；脱落
bring sth. to a halt	make sth. stop moving 使停止；使停住；使暂停
cut through	move or pass easily through water or air （在水中或空中）轻松地通过
fill up with	become full of sth.（使）充满；（使）装满
identify sb. as sb.	show who sb. is 表明…的身份
catch up with	come from behind and reach sb. in front of you by going faster 追上；赶上
break through (sth.)	manage to get past or through sth. that is in your way 冲破；突破
hold on	have your hands or arms tightly around sth. 紧紧抓住；紧紧抱住
let go (of sb. / sth.)	stop holding sb. or sth. 放手；松开
be grateful for	feel that you want to thank sb. because of sth. kind that they have done, or show this feeling 对…表示感激

- ***COME up with***

Phrases and expressions

excuse oneself	politely say that you are going to leave a place 礼貌地告辞
come up with	1 think of sth. such as an idea or an answer 想出，提出（主意或答案等） 2 produce or provide an mount of money 拿出，提供（钱款）
stick to sth.	1 do or keep doing what you said you would do or what you believe in 遵守；信守；坚持 2 keep using or doing one particular thing and not change to anything else 继续使用；继续做