

**THE PHRASEOLOGY OF PHRASAL VERBS IN  
ENGLISH: A CORPUS STUDY OF THE  
LANGUAGE OF CHINESE LEARNERS AND  
NATIVE ENGLISH WRITERS**

by

YUSHAN KE

A thesis submitted to the  
University of Birmingham  
for the degree of  
DOCTOR OF PHILOSOPHY

Department of English  
School of English, Drama and American & Canadian Studies  
College of Arts and Law  
The University of Birmingham  
September 2013

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## **ABSTRACT**

The aim of this study is to supplement existing research on phraseology in learner language by exploring the behaviours of phrasal verbs, a notorious hurdle for learners of English.

This thesis compares a Chinese learner corpus (CLEC) with an English native speakers' corpus (LOCNESS), with a reference corpus, the Bank of English (BoE), being consulted where necessary. A series of quantitative and qualitative investigations are conducted on phrasal verbs: calculation of frequency distribution and type-token ratios; identification of phraseological information, including collocation, semantic preference, semantic sequence and prosody. The results are discussed in full. Additionally, a framework utilising degrees of idiomaticity and restriction strength to group phrasal verbs is proposed and the issue of distinguishing synonymous counterparts is tackled as well.

The results generally indicate that Chinese learner language tends to have more phrasal verb tokens but fewer types than written native speaker English does. Detailed case studies of phrasal verbs show, however, that the phraseological behaviours of phrasal verbs as used by learners are so individualised that the findings are mixed. Learner uses are characterised by idiosyncrasies of different phraseological units, suggesting that the links (between lexis and lexis, or lexis and concepts) in the lexicon of L2 are different from those in L1.

## **DEDICATION**

*This work is dedicated to,*

*my beloved mother,*

*and my whole family*

## ***Acknowledgement***

First and foremost, praise and thanks to God for his blessings throughout my doctoral journey. This thesis would not have been possible without my supervisor, Professor Susan Hunston's, patient guidance. Her brilliant academic coaching, intellectual enlightenment and continuous support have led me along the path towards completing this work. I would like to express my sincere gratitude to her.

I owe thanks to many colleagues who have offered me help and support. Their kindness has made my life in Birmingham an enjoyable and rewarding experience. Among them, special thanks are due in particular to Dr. Makoto Sakai for his warm friendship. I am indebted to all of my beloved family members for their affection, sympathy and encouragement. Their unfailing love is the resource which fills me with the strength and enthusiasm to keep pursuing an academic career.

Thanks are also due to the National Science Council for providing me with a scholarship (NSC-095-SAF-I-564-058-TMS). Any error in this thesis is, of course, my sole responsibility.

This thesis was copy edited for conventions of language, spelling and grammar by Libro Editing Ltd.

## Table of Contents

THE PHRASEOLOGY OF PHRASAL VERBS IN ENGLISH: A CORPUS STUDY OF THE LANGUAGE OF CHINESE LEARNERS AND NATIVE ENGLISH WRITERS .....	1
Chapter1: INTRODUCTION .....	1
1.1 Introduction.....	1
1.2 Background of the thesis.....	1
1.2.1 Corpus, phraseological language and the contextual approach.....	1
1.2.2 Learner language features: unnaturalness.....	5
1.2.3 An overview of phrasal verbs .....	7
1.3 Motivation and problems .....	10
1.4 Aims and scope of the thesis.....	18
1.5 Guiding questions and potential significance of the thesis.....	19
1.6 Structure of the thesis.....	21
Chapter2: PHRASAL VERBS .....	24
2.1 Multi-word verbs .....	24
2.2 Definitions of phrasal verbs .....	25

2.3	Other relevant MWVs.....	29
2.3.1	Prepositional verbs and PVs .....	29
2.3.2	Free combinations and literal phrasal verbs .....	33
2.3.3	Phrasal-prepositional verbs.....	37
2.4	PVs in dictionaries .....	38
2.5	Elements, features and classification of phrasal verbs.....	40
2.5.1	Elements .....	41
2.5.2	Features and classification.....	42
2.6	Summary .....	58
Chapter3: LEARNER LANGUAGE, CORPORA AND PHRASAL VERBS		63
3.1	Introduction.....	63
3.2	Learner language.....	64
3.2.1	LL characteristics.....	64
3.2.2	Learners' lexical knowledge .....	65
3.2.3	Development of learner language studies.....	67
3.3	Corpora and teaching/learning .....	72

3.3.1	Corpora and teaching.....	72
3.3.2	Phraseology and teaching.....	76
3.4	Learning PVs .....	77
3.4.1	Collocation and PVs .....	77
3.4.2	Problems of learning and teaching PVs.....	79
3.4.3	Other studies of PVs .....	84
3.5	Summary .....	86
Chapter4:	PHRASEOLOGY .....	88
4.1	Introduction.....	88
4.2	Phraseology and other related concepts.....	90
4.2.1	Phraseology .....	90
4.2.2	Idiomaticity.....	92
4.2.3	Collocation.....	93
4.3	Phraseology and meaning .....	102
4.3.1	Units of meaning .....	103
4.3.2	Semantic/pragmatic associations and lexical priming.....	107

4.3.3	Semantic sequence.....	110
4.3.4	Contextual approach.....	114
4.3.5	Phraseology/collocation in learner language.....	115
4.4	Summary.....	123
Chapter5:	METHODOLOGY.....	126
5.1	Preliminaries.....	126
5.2	Corpora and tools.....	130
5.2.1	Corpora selection and structure.....	130
5.2.2	Software.....	136
5.2.3	Materials.....	138
5.3	Procedure.....	138
5.3.1	Extraction of PVs.....	138
5.3.2	Extraction of phraseological units: collocations, associations and sequences	153
5.4	Terminology.....	161
5.5	Summary.....	162

Chapter6:	PHRASAL VERBS WITH THE PARTICLE ‘UP’ .....	164
6.1	Introduction.....	164
6.2	Verb types, TTR and over/under-representation .....	167
6.3	Analyses of five example PVs .....	173
6.3.1	The Chinese learner language performance: the five example PVs 175	
6.4	A visual illustration of PVs .....	190
6.4.1	The ambiguity of idiomaticity and collocation restriction .....	190
6.4.2	Bi-axis illustration of PVs .....	198
6.5	Summary .....	218
Chapter7:	PHRASAL VERBS WITH THE PARTICLE ‘OUT’ .....	221
7.1	Introduction.....	221
7.2	Overall results .....	222
7.3	The case of CARRY OUT .....	224
7.3.1	Lexico-grammatical analysis .....	225
7.4	The case of FIND OUT.....	241

7.4.1	Introduction .....	241
7.4.2	Usage patterns of FIND OUT in CLEC, LOCNESS and BoE.....	242
7.4.3	Comparing FIND and FIND OUT in CLEC and BoE .....	258
7.5	Summary .....	277
Chapter8: PHRASAL VERBS WITH PARTICLES ‘ON’, ‘ABOUT’ AND		
‘DOWN’	281	
8.1	Introduction.....	281
8.2	PVs with ‘ON’ .....	283
8.2.1	GO ON.....	284
8.2.2	TAKE ON .....	299
8.3	PVs with ‘ABOUT’ .....	305
8.3.1	BRING ABOUT .....	305
8.3.2	COME ABOUT .....	309
8.4	PVs with ‘DOWN’ .....	312
8.4.1	BREAK DOWN .....	313
8.4.2	CUT DOWN.....	317

8.5	Summary .....	321
Chapter9:	DISCUSSION AND PEDAGOGICAL IMPLICATIONS .....	325
9.1	Introduction.....	325
9.2	Review of major findings.....	326
9.2.1	Question 1: The distributional features of PVs .....	326
9.2.2	Question 2: Idiomaticity and restriction strength of PVs .....	335
9.2.3	Question 3: Discriminating a PV and its synonym.....	339
9.2.4	Question 4: The Chinese Learners' uses of PVs, with a focus on phraseologies.....	340
9.3	Discussion .....	346
9.3.1	Influence of topics, genres or registers .....	347
9.3.2	L1 influence .....	349
9.3.3	L2 lexicon .....	355
9.4	Pedagogical recommendations.....	361
9.4.1	Classing PVs is not absolutely necessary .....	362
9.4.2	PVs with an identical particle can better be presented together, in	

texts of specific topics, notions, genres or registers.....	363
9.4.3    Selecting appropriate PVs for teaching .....	365
9.4.4    The two-dimensional model of PVs and its implications for teaching order.....	367
9.4.5    Contrasting PVs in conjunction with the synonymous one-word verb	371
9.4.6    Focusing on the phraseologies of PVs.....	372
9.5    Summary .....	376
Chapter10: CONCLUSION .....	378
10.1    Introduction .....	378
10.2    Review of the thesis .....	378
10.3    Limitations .....	384
10.4    Future avenues.....	387
10.4.1    A word of caution for future research .....	387
10.4.2    Directions for future research.....	389
10.4.3    The challenges of learning phraseologies .....	391

10.5	Concluding remarks .....	394
------	--------------------------	-----

## List of Figures

Figure 3.1: The example of “draw up guidelines” .....	77
Figure 5.1: A screenshot of the original data with error tags and identity annotations .....	137
Figure 5.2: A screenshot of the data cleaned by PowerGREP .....	137
Figure 6.1: The two-dimensional model of PVs .....	215
Figure 7.1: Proportional distribution of patterns .....	247
Figure 8.1: Patterns of GO ON captured by WordSmith4 in CLEC .....	286
Figure 8.2: Patterns of GO ON captured by WordSmith4 in LOCNESS .....	286
Figure 8.3: Instances of ‘ <i>as time</i> GO ON’ in CLEC .....	290
Figure 8.4: Examples of ‘what is going on’ and ‘what goes on’ .....	298
Figure 8.5: Example patterns of TAKE ON in CLEC .....	300
Figure 8.6: Example patterns of TAKE ON in LOCNESS .....	300
Figure 8.7: Examples of ‘TAKE ON + a +Adj +look/job’ .....	302
Figure 8.8: Concordance lines of TAKE ON in CLEC .....	303
Figure 8.9: Concordance lines of TAKE ON in LOCNESS .....	304
Figure 8.10: Patterns of BRING ABOUT in CLEC .....	305
Figure 8.11: Patterns of BRING ABOUT in LOCNESS .....	306
Figure 8.12: Patterns of BREAK DOWN in CLEC .....	313

Figure 8.13: Patterns of BREAK DOWN in LOCNESS.....	313
Figure 8.14: Examples of BREAK DOWN from CLEC.....	314
Figure 8.15: Examples of BREAK DOWN in LOCNESS.....	315
Figure 8.16: Examples of BREAK DOWN with ‘items of bad quality’ in CLEC .....	316
Figure 8.17: Examples of BREAK DOWN with ‘computer/car’ in LOCNESS	317
Figure 8.18: Examples of CUT DOWN in CLEC .....	317
Figure 8.19: Examples of CUT DOWN in LOCNESS.....	317
Figure 8.20: Concordance of CUT DOWN ON from LOCNESS.....	318
Figure 8.21: KWIC of REDUCE from CLEC .....	320

## List of Tables

Table 2.1: Types and examples of multi-word verbs .....	25
Table 2.2: Survey of the inclusion of PVs in research and dictionaries (N=No, Y=Yes) .....	40
Table 2.3: Examples of particles .....	41
Table 2.4: Classification of PVs in previous studies.....	48
Table 2.5: Properties of different VPCs .....	61
Table 3.1: Most productive verbs and particles of phrasal verbs.....	85
Table 4.1: Examples of semantic sequences .....	112
Table 5.1: Topics in CLEC.....	131
Table 5.2: Structure and size of CLEC .....	133
Table 5.3: Structure and size of LOCNESS.....	133
Table 5.4: T-score vs. MI-score.....	158
Table 6.1: The top PVs in CLEC and LOCNESS.....	169
Table 6.2: The top five over-/under-represented PVs in CLEC and LOCNESS	171
Table 6.3: Attributes of the five phrasal verbs .....	175
Table 6.4: Collocation of the five selected PVs in CLEC.....	176
Table 6.5: Collocation of the five selected PVs in LOCNESS .....	176
Table 6.6: Functions of linguistic items around BRING UP .....	183

Table 6.7: Selecting targets among the top20 collocates of GROW UP in BoE (Span 4:4).....	205
Table 6.8: Typical collocates of five example PVs.....	208
Table 7.1: Top five most frequent verbs in CLEC and LOCNESS.....	223
Table 7.2: Frequencies of word forms of CARRY OUT .....	226
Table 7.3: Occurrences of agents and patients/themes across corpora.....	228
Table 7.4: Proportional distribution of ‘concepts’ with ‘ <i>carried out</i> ’ .....	239
Table 7.5: Distribution of following patterns of ‘FIND OUT’ in CLEC and LOCNESS.....	243
Table 7.6: The top 15 most frequent collocates of FIND OUT in BoE.....	251
Table 7.7: Frequencies and percentages of some patterns in BoE.....	252
Table 7.8: Frequencies and percentages of the six collocates with FIND OUT in BoE and CLEC .....	254
Table 7.9: List of words in the EVALUATION slot .....	257
Table 7.10: Forms of FIND in BoE .....	263
Table 7.11: Forms of FIND OUT in BoE .....	263
Table 7.12: Patterns of FIND OUT in BoE by frequency.....	264
Table 7.13: Patterns of the word form ‘find out’ in BoE by frequency .....	265

Table 7.14: Patterns of the word-form ‘found out’ in BoE by frequency .....	266
Table 7.15: Patterns of the word-form ‘found out’ in BoE by frequency .....	267
Table 7.16: Patterns of the word-form ‘find’ in BoE by frequency .....	268
Table 7.17: Patterns of the word-form ‘found’ in BoE by frequency .....	269
Table 7.18: Summary of typical patterns of FIND OUT and FIND .....	272
Table 8.1: Patterns of GO ON .....	287
Table 8.2: Patterns of TAKE ON .....	300
Table 8.3: Patterns of BRING ABOUT .....	306
Table 9.1: Normalised frequencies of PVs across five particle groups .....	328
Table 9.2: Type-token ratios (%) across the PV groups.....	329
Table 9.3: The five most frequent PV types across groups in LOCNESS and CLEC .....	332
Table 9.4: The most frequent PV types of G-ICLE and I-ICLE (Waibel, 2007:87) .....	333
Table 9.5: PVs’ different emphases by the Chinese learners and native students .....	345

## List of conventions

1. “ ” **Double quotation marks:** for short direct quotations from the original publication and for marking example phrases or sentences
2. ‘ ’ **Single quotation marks:** for terminologies, senses or meanings, and my emphases
3. [ ] **square brackets:** for semantic fields/associations or concepts (in semantic sequences)
4. **UPPER CASE:** for abbreviations, and for indicating the lemmas of the PVs analysed in this thesis
5. *Italics:* for examples and instances, and book names

# **Chapter1: INTRODUCTION**

## **1.1 Introduction**

This thesis studies phrasal verbs in one corpus of learner language (CLEC) and one of native language (LOCNESS), from the perspective that the usages of a phrasal verb can be profiled with its '(extended) units of meaning' (Sinclair, 1996). It is motivated by the pedagogical needs that phrasal verbs present to foreign language learners and the advance of the study of phraseologies in corpus research.

This chapter will first set the background of learner language (abbreviated as LL hereafter) characteristics, the phraseological nature of language and the properties of phrasal verbs (PVs, henceforth). Potential problems will be identified with the research questions which are to be addressed. The scope of this work is then reported, followed by a discussion of the potential contributions of this research. This chapter ends with a summary of the outline of this thesis.

## **1.2 Background of the thesis**

### **1.2.1 Corpus, phraseological language and the contextual approach**

A corpus is a large collection of texts which can be researched by linguists with the assistance of computers. Since the 1960s, corpora have been made digital and able to

be processed by computer programs. Thanks to advances in modern technology, personal computers are powerful enough to enable linguists to deal with a large amount of data. With the advantages of efficiency and huge data-processing ability, the corpus approach has become popular in the field of applied linguistics, for instance, lexical studies, studies of register, translation studies, and comparisons between languages (McEnery, Xiao, & Tono, 2006:80-122).

One significant finding resulting from the corpus method is that phraseology lies at the heart of language description. The core of language has been shown to consist largely of many prefabricated constructions or ready-made expressions and patterns, which will be collectively called phraseologies. As the importance of phraseology has been recognised by linguists, there is a burgeoning literature studying this phenomenon. A large number of studies have been conducted on phraseologies with different names such as 'phraseological units', 'prefabs', 'formulaic expressions', 'chunks', 'lexical bundles', 'collocation', 'multi-word units', etc. (e.g. Granger, 2005; Wray, 2002, see Chapter 4). All of these are derived from the fact that language is full of combinations of these lexis-based constructions. In the profession of language studies, the term 'phraseology' usually refers to the clustering of words, but it can also be extended to

indicate various co-occurring linguistic phenomena (see Chapter 4 for details), because language is found to be constrained by many co-occurrence restrictions.

Before we discuss these restrictions, attention should be focused on an equally important discovery of corpus studies: the drastic change of our perspective on word meaning. With the support of concrete corpus evidence, Sinclair (2004b:25) points out that “words enter into meaningful relations with other words around them”. The long-standing belief that the meaning of a word can stand alone and is separable from the context has been challenged. The meaning takes its form when other words join together; in other words, the meaning or sense is shaped and defined by what co-occurs around it. The meaning is created by accumulation of words, i.e. word-word collocation, which has been the main focus in many studies, for example: Stubbs (1995), Howarth (1996), Nesselhauf (2003, 2005), Lennon (2005), Lesniewska (2006), etc. Besides word-word collocation, the phenomenon of co-occurrence can also be found at other levels. Sinclair (2004b:24) describes all the levels which contribute to the meaning as ‘extended units of meaning’, including collocation, colligation, semantic preference and semantic prosody (cf. Section 4.3.1).

Some researchers have looked at targets of scope larger than words; for instance, Hoey (2005) proposes and attests the idea of ‘semantic association’, which confines the

possible presence of words, and Hunston (2008) introduces ‘semantic sequence’, which shows patterns of “series of meaning elements” in languages (see Section 4.3.3). The evidence of these linguistic phenomena shows that language is controlled not only by grammatical rules but also by lexical and discursal co-selections. Many tacit factors in the context will place restrictions on the co-occurrences of words. The restrictions of co-occurrences at different linguistic levels result in the ‘un-randomness’ of language.

The view that language is restricted at several levels exerts great influence on the way in which language is described. As discussed, attention has been drawn to not only the relations between lexical items, but also their relations to the factors of context. All of the aspects of a lexical item and its context constitute its usages. The importance of examining language usages is emphasised in Barlow and Kemmer (1999), where various language study approaches based on usages are brought together. The corpus approach certainly represents one of these usage-based approaches, because it looks at authentic data produced by language users. With a focus on word usages, Biber et al. (1999:289-290) addresses the significance of investigating ‘association patterns’, which are systematic co-occurrences including both lexical and grammatical associations, for example, the collocations of a particular word (i.e. lexical associations) and the structural preferences (i.e. grammatical associations). These ‘association

patterns' and the aforementioned 'extended units of meaning' and other patterned expressions such as 'semantic associations' (Hoey, 2005) and 'semantic sequences' (Hunston, 2008) all contribute to describe language usages. They place restrictions and constraints on languages, keeping those languages consistent, systematic, and phraseological. The phraseologies of languages can thus be taken as the means for identifying the characterisations of different language types, such as native language and learner language.

### **1.2.2 Learner language features: unnaturalness**

Many branches of language studies benefit from the application of corpora, the study of learner language being one such area which has gained new insights by using corpora. Looking at learner language through a large collection of texts can reveal the particular characteristics of this special type of language. Learner language has been depicted as "informal, speech-like" (Granger & Rayson, 1998:130), "bookish and pedantic" (Channell, 1994:21), "vague and stereotyped" and having "limited vocabulary" (Ringbom, 1998:49), or lacking idiomaticity (Lorenz, 1998:53). All of these suggest that learner language has its own style, which is generally referred to as 'unnatural' or 'non-native'.

The ‘unnaturalness’ of learner language may be illustrated by misuses of words, grammatical errors, inappropriate choice of vocabulary, etc. These kinds of ‘local error’ are easily detected, but there are more ‘global errors’ which extend through larger scopes (Burt & Kiparsky, 1972), such as improper co-selection or odd combinations of words and sentence patterns, as pointed out by Shei (2005:218). He comments that a seemingly non-native-like sentence produced by a learner “may have to be rephrased, choosing the appropriate lexical units and their habitual sentence pattern to express the desired meaning”. This remark shows that learners face huge challenges in combining lexical elements to form native-like expressions, in particular phraseologies.

Such unnaturalness is difficult to capture using conventional approaches such as error analysis. However, with a corpus-driven approach, the gaps between learner language and NS language can be revealed from a new perspective. With corpora of a good size and specialised software, phraseologies could be effectively brought to light through observing the differences of repetitive patterns. By comparing both native and non-native corpora, a number of characteristics of learner languages have been successfully discovered in studies such as De Cock et al. (1998), Lorenz (1998) and many others, as will be seen in Chapter 3. Therefore contrasting native and learner

corpora is believed to be a fruitful area from which researchers can extract important information of learners' un-naturalness.

### **1.2.3 An overview of phrasal verbs**

When corpus studies created a surge of interest, one special group, phrasal verbs, drew researchers' attention immediately. Although the study of PVs is one of the most remarkable targets in the study of phraseology, they are so distinct from other phraseological targets that they need "separate and thorough research of their own" (Grant & Bauer, 2004:39).

A phrasal verb consists of a verb with a particle. This construction has been given various names by different researchers, such as verb-particle constructions, multi-word verbs, compound verbs, complex verbs, particle verbs, composite verbal expressions, discontinuous verbs, etc. (e.g. Lam, 2003; Schneider, 2004, see Chapter 2), and they have also been variously defined by different researchers (Claridge, 2000). For example, Claridge (2000) conceives phrasal verbs as a subtype of multi-word verbs, while other researchers have interpreted PVs as the combinations of "a head verb and one or more obligatory particles, in the form of intransitive prepositions, adjectives or verbs" (Baldwin & Villavicencio, 2002:98): examples such as *hand in*, *cut short*, *let go* are given in their paper. Generally, a PV is defined as "a structure that consists of a verb

proper and a morphologically invariable particle that function as a single unit both lexically and syntactically” (Liao & Fukuya, 2004:196). Or it is regarded as “idiomatic combinations of a verb and one or more particles which jointly behave as a single lexico-grammatical unit” (Schneider, 2004:230). The common description of PVs in these definitions is the unity of the verb and its particles.

Besides the particular definition of the united construction of PVs, various characteristics can be applied to some but not all phrasal verbs: (1) Some of them are polysemous, e.g. *make up* has eight senses (Villavicencio, 2003). (2) Their syntactic patterns can be flexible or fixed, intransitive or transitive, or even di-transitive. (3) The existence of the particle can be omissible or obligatory, for example: *wake (up)* but *get up*. (4) Their meanings can be of different degrees of idiomaticity, from literal to figurative or idiomatic, e.g. *stand up* can mean ‘rise from a sitting/lying position’ or ‘an idea is proved to be correct’.

Several researchers have tried to divide PVs into different categories, for example: (1) literal: *go out, take away, come in* (2) figurative: *turn up, let down* (3) completive: *cut off, burn down* (Dagut & Laufer, 1985). Also, PVs can be classified as semantically transparent, semi-transparent, and figurative/ semantically opaque according to their semantic transparency (Laufer & Eliasson, 1993). Or they could be simply classified

into compositional, idiomatic or aspectual, according to their semantic interpretations (Dehé, Jackendoff, McIntyre, & Urban, 2002; Dehé, 2002). The categorisation of PVs is indeterminate, which is a consequence of their semantic complexities.

The above-mentioned complexities of phrasal verbs have been observed to create great challenges for learners from different language backgrounds. For example, researchers such as Siyanova and Schmitt (2007:120) have remarked that PVs are rather problematic for non-Germanic or non-Scandinavian students learning English as a foreign language. In particular, figurative PVs have been reported as being confusing for Chinese learners (Liao & Fukuya, 2004:215). The difficulties arise from many factors. The multiplicity of senses of PVs is recognised as a major hurdle for L2 learners by Cornell (1985). Side (1990:144-145) gives a full summary of the difficulties learners may confront (e.g. “confusion of combining the verb and the particle” and “polysemy”, see Section 3.4.2 for details). Lennon (1996) further adds “contextual and collocational restrictions” and “grammatical environment”, which have been considered marginal in the previous literature, but are of particular interest to this present study.

The studies conducted in light of learners’ difficulty with respect to PVs have revealed several general findings (Hägglund, 2001; Sjöholm, 1995). Firstly, the

structural difference of L1-L2 will confound learners. Secondly, literal PVs are widely preferred to idiomatic PVs across different L1 backgrounds, and idiomatic PVs are often regarded as the most difficult type for ESL/EFL learners (Celce-Murcia & Larsen-Freeman, 1999:274). Thirdly, learners tend to use PVs less frequently than native speakers. A group of studies have focused on the phenomenon of ‘avoidance’ of PVs (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004). Therefore, PVs are a special group of linguistic items which learners often deem problematic, displaying reluctance to use them.

### **1.3 Motivation and problems**

This study is motivated by the desire to help Chinese students write natural-sounding English in a foreign language environment. One of the challenges of learning English for Chinese learners is native-like production, which even the most advanced learners rarely achieve. With the development of the Computer Learner Corpus (CLC), it has become more convenient and easy to detect these non-native-like features through comparing native-speakers’ (NSs) and non-native speakers’ (NNSs) corpora. Even though CLC studies indeed facilitate the investigation of non-nativeness, the results from previous CLC research have not been wholly satisfactory.

This is partly because of the traditional perspective of teaching and learning English. In the past, teachers often only corrected grammatical errors for students, leading the English teaching approach to focus on grammar. Previous work on describing the deviances of learner language in CLC research has also mainly concentrated on comparing frequencies of uses (i.e. overuses and underuses) in the learner language and target language corpora, or stating errors through error-tagged corpora (Diaz-Negrillo & Fernandez-Dominguez, 2006:84). However, these methodologies cover only part of the language repertoire of learners, i.e. frequency and errors. James (1998:65-70) explains that errors can be identified for breach of either ‘grammaticality’ or ‘acceptability’ (or both). The former indicates grammatical, semantic and phonological well-formedness and the latter refers to contextual appropriacy (for example, “*She decided to answer the telephone call*” is unacceptable when the speaker intends to ‘pick up the receiver while it is ringing’, see James (1998:68)). As the intention of the speaker/writer is usually not clear, errors in corpus study are mainly limited to the former type. Therefore we need a broader view in describing and explaining learner language as a unique system, which should not be envisaged as only confined to negative aspects but should be described in a comprehensive view. In order to help learners achieve ‘nativeness’, we should also pay attention to the areas where no errors

are shown but there are detectable differences which display the features of non-nativeness.

To compensate for the weakness of traditional error analysis, some researchers in learner corpus studies have advocated a new way of research. In her much-quoted study, Granger (1998a:13) proposes Contrastive Interlanguage Analysis (CIA), which states the importance of comparing the languages of NSs and NNSs. Granger also reminds us that the purpose of a learner corpus is to “uncover the features of non-nativeness of learner language”, with an emphasis that the features should not only be focused on “plain errors”, but should also cover the “frequency of use of certain words, phrases or structures”. Leech (1998:xvii) also points out that learner corpus research enables us to explore not only what the learners did wrong but also what they did right. This is true and essential for describing learner languages, and we can gain a great deal of information from those expressions which do not contain errors. Even if a learner does not make any errors, their wording may still seem to be different from native writing, as observed by Shei (2005) (c.f. Section 1.2.2). The general impression of learner language is conceived as less expressive in contrast to products from native writers. This cannot be fully accounted for by simply judging whether the learners can write correct sentences. We need a new view and approach to access learners’ language.

An all-inclusive perspective can be underpinned by the theories which arise from the studies of co-occurrences in language. Such research has attested to the existence of ‘phraseology/idiomaticity’ in language, such as idioms, collocation, phraseological units, semantic association, sense-structure patterns, textual fixedness, etc. (Hoey, 2005; Howarth, 1998; Hunston & Francis, 1999; Moon, 1998; Nesselhauf, 2003; Shei, 2005; Sinclair, 1991, among many others). These studies have shed light on how language can be described holistically. The increasing numbers of such studies suggest that the importance of phraseology is mounting. Therefore, co-occurrences at several linguistic levels, such as lexical or grammatical associations, etc., which are termed ‘extended lexical units’ by Sinclair (2004b), will be the major concerns of this research. The term ‘phraseology’ will be adopted in a wide sense to account for all of these relevant issues.

That a large number of language constructions are fixed, prefabricated or idiomatic raises the question as to whether learner language also has similar phenomena. Sinclair (1991:110) proposes the ‘idiom principle’, which states that language users have many ‘semi-preconstructed phrases’ on hand, and other studies which have looked at phraseologies have also substantiated that LL is phraseological to a certain extent as well (see Chapters 3 and 4). By investigating how learner language differs from NS

language in combining or sequencing lexical items, this area can fruitfully contribute to our understanding of learner language.

This thesis will focus on phrasal verbs, which are used as an example by Sinclair (2004a:26) that casts doubt on the assumption that ‘words are independent in a language’. Many other studies of formulaic languages also concentrate on PVs (see Chapter 2), demonstrating that PVs are the fertile field *par excellence* to explore phraseologies. The second reason lies in the difficulties PVs present to learners in English learning. PVs are always a hot issue raised in a foreign language setting, and are often treated independently in textbooks, because their behaviours are so particular and complicated. They have been regarded as a thorny problem for the linguistic complexities they carry.

Earlier, we have seen that PVs have complicated features such as being polysemous, idiomatic, etc., which often cause stumbling blocks for learners. The reason that PVs are worthy of meticulous attention can be addressed by their difficulty to learners. Difficulties with PVs may stem from structure divergence across languages. It is reasonable to assume that Chinese English learners will have problems acquiring this particular language structure. In fact, PVs have been demonstrated to result in problems for learners, for example, Dagut and Laufer (1985), De Cock (2005) (see Section 3.4.2).

As this research studies the language of Chinese learners, their first language (Chinese) is bound to interfere with their second language performance. PVs appear to cause serious problems, especially to Chinese learners, because PVs in Chinese and English have largely dissimilar properties. Although Chinese has PVs, they are different from those in English in that the particles are inseparable from the verbs, and there are fewer particles; moreover, PVs in Chinese rarely have figurative meanings (Liao & Fukuya, 2004:211). All of these mentioned above have contributed to the difficulties of learning PVs, posing the need to investigate this acute problem. Given these impetus, the focus of this study will thus be placed on phrasal verbs.

The discovery that the meaning of a linguistic item has consonance with its associated phraseologies has drawn many researchers' interest to the contextual characteristics; thereby the research direction will be steered towards discovering these phraseological associations. Many types of linguistic item, such as verbs with nouns, adverbs, and some discourse features, have been investigated in terms of their phraseologies. These studies have made substantial contributions to our knowledge about language. Unfortunately such an attempt has not yet been applied to PVs, which are a significant area at the core of idiomaticity/phraseology studies, and a great challenge for Chinese learners. Studies on PVs have set their eyes on mismatches of the

verb and its particle or problems in respect of PVs' syntactic complexity and semantic opaqueness (see Chapter 2). Their concerns are limited to the PVs themselves without considering co-occurring factors. Outside factors (semantic association/ prosody/ sequence), although they have been proved vital, have not been attended to in the area of PVs. Little attention has been paid to their actual usages, which are defined by the PV itself and its context. To take a wider view of the behaviours of PVs, this thesis advances the scope to explore more phraseological elements which distinguish the Chinese learner language and the native English language.

In pedagogy, researchers adopting a cognitive approach have often suggested that the best way to learn PVs is to grasp the uniformed patterns of the particles, and generalise from the fundamental sense when encountering new PVs (see Section 3.4.3). The pedagogical focus has been concentrated on the entangled syntactic and semantic features of PVs. These research tendencies are admittedly helpful in embodying knowledge in respect of PVs. However, these pedagogical approaches have some limitations. For instance, the cognitive approach can save learners' efforts as they only need to learn the basic meanings of the particles, but they will still fall short of employing a particular PV at their disposal because the cognitive analysis of the particle will only function in the receptive but not the productive learning process.

Teaching learners the complexities of PVs will result in the learners becoming very confused. One solution may be to pay attention to the usages of PVs in context, but such a study unfortunately has not been conducted so far.

The most vital gap in our exploration of LL has been pointed out by Cowie and Howarth (1996:88): “little is known in detail about phraseological competence in a second language, nor about how it is acquired”. A few decades have passed, but progress in this field is still advancing rather slowly. Not all phraseological units are given the same consideration. Researchers of LL tend to be interested in studying certain phraseological phenomena such as prefabricated sequences/formulae (e.g. De Cock, Granger, Leech, & McEnery, 1998; Granger, 1998b) and collocations (Handl, 2008; Howarth, 1996; Lennon, 2005; Nesselhauf, 2005). Other phraseological phenomena such as semantic preference and semantic prosody, even though their importance has been well established in native English (e.g. Kennedy, 2008; Louw, 1993; Partington, 2004), have not drawn as much attention in the domain of learner language studies. Taken together, a study which considers comprehensive factors (phraseological behaviours) in relation to one specific linguistic group is missing in the literature; therefore this thesis hopes to throw new light on learner language research.

## **1.4 Aims and scope of the thesis**

The premise assumed in this study is that the native speakers of a language have subconscious knowledge or intuition about the uses of their mother tongue, which is manipulated by certain implicit habitual restrictions. However, these conventions are not shared by non-native language learners to the same extent. Therefore, the main thesis of this present research is that language learners use the target language very differently in contrast to native speakers, not only in terms of frequencies, but also in the manifestation of word selection and combination. This research will describe the different uses of PVs between Chinese English learners and English native speakers in terms of individual words and associated phraseologies, based on a corpus-driven approach. Non-native-like characteristics regarding PVs in a Chinese learner corpus will be reported both quantitatively and qualitatively. Moreover, this study also serves to provide some pedagogical recommendations for phrasal verb teaching or learning, in order to apply the findings to practice.

The research target of this study is PVs, which are two-word constructions consisting of a verb and a particle. For manageability, the particles under examination will be restricted to five targets, i.e. UP, OUT, ON, ABOUT, DOWN. These particles are randomly chosen (see Chapter 5). PVs with these particles will be examined

quantitatively, and some PVs will be further examined qualitatively in light of their phraseological aspects. These intensively probed PVs include (1) V+UP group: DRAW UP, LOOK UP, BRING UP, GROW UP, PICK UP; (2) V+OUT group: CARRY OUT, FIND OUT; (3) V+ON group: GO ON, TAKE ON; (4) V+ABOUT group: BRING ABOUT, COME ABOUT; (5) V+DOWN group: BREAK DOWN, CUT DOWN. The lexical associations (e.g. collocates, semantic fields), grammatical associations (e.g. word-form preferences, structure preferences) and combinatorial associations (e.g. semantic sequences) of these examples will be studied. In addition, the question in respect to the degrees of restriction and idiomaticity will be answered using the example of the V+UP group. A comparison of FIND and FIND OUT will also be made to illustrate usage differences between apparent synonyms.

## **1.5 Guiding questions and potential significance of the thesis**

As discussed above, Chinese learners encounter considerable difficulties in using PVs and will be beset with problems in attempting to achieve native-likeness. This problem of unnaturalness can be approached at two levels: the lexical level and the contextual level. At the lexical level, general profiles for the PVs will be provided from the angles of distribution, type-token ratios, verb/particle productivity,

idiomaticity and preference for word-forms of one lemma. At the contextual level, a number of phraseological associations will be examined, e.g. collocations, semantic and grammatical patterns. Taken together, the uses of a PV will be compared in all aspects to examine the Chinese learners' special style, which bears little resemblance to native English, if nativeness is taken as the goal of learning.

This thesis concerns a main question that can be put: **How do the usages of phrasal verbs differ in native English and Chinese learner language?**

In order to answer this chief question, a number of more specific sub-questions will be addressed:

1. How do the Chinese learners (CLEC) and native English writers (LOCNESS) use PVs differently in terms of distribution (e.g. frequency of occurrences, type-token ratios and the most frequent PV types)?
2. How do the degrees of idiomaticity and restriction strength help to characterise PVs, based on data from an English reference corpus (BoE)? (Chapter 6)
3. How can a phrasal verb be distinguished from its near-synonym, in the Chinese learner corpus (CLEC) and the English native corpora (LOCNESS and BoE)? (Chapter 7)

4. How do the Chinese learner (CLEC) uses of PVs differ from native uses (LOCNESS) in terms of phraseological units? (Chapters 6 to 8)

As indicated by the research questions, this study will delineate the differences of using PVs by the Chinese learners as compared to native English writers. A comparison of LL to native speakers' language is intriguing in that the preference for certain 'patterns' can be explored and the area of 'usages' which is neglected by language teaching, can be supplemented. It is envisaged in this study that the Chinese learners may not possess full knowledge of the phraseologies associated with a PV, resulting in non-native-like characteristics. Highlighting the usage differences with a contextual approach is believed to shed light on our understanding of the roles of phraseologies in the Chinese learner languages. Besides the theoretical implications, the findings of this thesis are also hoped to inform teaching methods, in order to alleviate the burden of learning a complex construction in a foreign language.

## **1.6 Structure of the thesis**

This thesis is divided into nine chapters. Chapter 1 introduces the background, aims and scope of the study. The second chapter provides an outline of several sub-types of multi-word verbs. The definitions and characteristics of phrasal verbs are then

presented, followed by a detailed explanation of the semantic and syntactic complexities of PVs. Chapter 3 reviews the general description of learner language, and discusses corpus approaches in terms of their advantages and deficiencies. The theoretical framework of lexical knowledge is articulated, and the applications of corpora in teaching are covered. In addition, issues of learning and teaching phrasal verbs are presented, in order to set in context the kinds of problem that learners may encounter and how PVs are treated in pedagogical environments. The problems of PVs in respect to collocation are particularly discussed, with the empirical findings from previous studies summarised. The other important issue, phraseology, is the topic of Chapter 4. This chapter discusses the ideas involved with phraseology, i.e. the versatility of phraseology. The approaches to extracting phraseological units are illustrated, and a number of co-occurrences at different levels are reported. This chapter also takes the readers through research examining the roles of phraseology and collocation in the learner language. The fifth chapter elaborates the methods in relation to corpus selection and data extraction. The size and structure of the corpora are discussed and the measures taken to ensure comparability described. The reasons for the selection of material and the design of the procedures are justified. The extraction of the PVs and their phraseological units are explicated step by step.

The next three chapters (6-8) each begin by presenting numerical data, and then broaden the analysis to cover contextual features. Some specific questions which arise from the research findings are tackled in the chapters respectively. Chapter 6 deals with V+UP constructions. An illustration of PVs by their degrees of idiomaticity and restrictions is proposed, and five selected PVs are investigated in terms of their collocations. Chapter 7 illustrates the behaviours of V+OUT constructions, focusing on two PVs: CARRY OUT and FIND OUT. More attention is paid to their contextual features, and a comparison of the synonyms FIND and FIND OUT is made. The following Chapter 8 broadens the research targets to cover constructions with less frequent particles, such as V+ON, V+ABOUT and V+DOWN. The main findings are summarised and the pedagogical implications discussed in Chapter 9. The final chapter, Chapter 10, concludes the contributions and provides directions for future research.

## Chapter2: PHRASAL VERBS

### 2.1 Multi-word verbs

With their distinctive linguistic behaviours, multi-word verbs (MWVs) have received substantial attention in previous studies. In fact, MWVs consist of many different subcategories, and the category of phrasal verbs is usually included within the broad concept of multi-word verbs. Quirk et al. (1985:1150) classify MWVs into three major categories: phrasal verbs (e.g. *turn up*), prepositional verbs (e.g. *dispose of*) and phrasal-prepositional verbs (e.g. *get away with*). In this definition, a PV is made of ‘V+adverb’, a prepositional verb consists of ‘V+preposition’, and a phrasal-prepositional verb contains ‘V+adverb+preposition’. As these three groups are taken as MWVs, they all behave like single-unit verbs, and the non-verbal parts in these three groups, termed ‘particles’ by Quirk et al., will be used as a neutral term which does not indicate its part-of-speech role throughout this thesis. Besides these three categories, other constructions are found, such as verb-adjective combinations (e.g. *hold good*), verbo-nominal combinations (e.g. *put in execution*) (Claridge, 2000), and some marginal types such as verb-verb combinations (e.g. *make do with*), verbs governing two prepositions (e.g. *develop ... from ... into*) (Quirk et al., 1985:1167-1168). Examples of these MWVs are given in Table 2.1. Among them,

phrasal verb (PV) is certainly the most familiar term for language learners and teachers.

A brief introduction of the linguistic properties of phrasal verbs will now be presented,

followed by a discussion of some issues which are examined in my thesis.

**Table 2.1: Types and examples of multi-word verbs**

MWV Types	Subtypes	Examples
Phrasal verbs	(a) intransitive	<i>Blow up, get on</i>
	(b) transitive	<i>Set up, put off</i>
Prepositional verbs	(a) Type I	<i>Cope with, rely on, look at, approve of</i>
	(b) Type II	<i>Confine NP to, protect NP from</i>
Phrasal-prepositional verbs		<i>Come up with, look forward to</i>
Verb-adjective combinations		<i>Put NP straight, lay (NP) low</i>
Verbo-nominal combinations		<i>Take place, set fire to, bring to light</i>
Verb-verb combinations		<i>Make do with, let NP go</i>
Verbs governing two prepositions		<i>Develop from NP into NP, talk to NP about NP</i>

## 2.2 Definitions of phrasal verbs

PVs are one type of MWV with unique characteristics which separate them as a group that is different from other verbs. However, the label ‘phrasal verbs’ is a problematic term, because it has been polysemous and multi-purpose in the literature. In addition, the definition of PVs has not been consistent, due to their complicated nature and different research purposes found in the literature.

In terms of nomenclature, PVs are sometimes called ‘verb-particle constructions

(VPCs)' or 'particle verbs' (Schneider, 2004:229). There is little agreement in the literature on terminologies such as VPC and PV. They have been defined differently across research studies, resulting in considerable confusion (Dalle, 1983:12; Lam, 2003:76). For instance, the construction made of a verb and a particle is termed a verb-particle construction (VPC) by researchers such as Lam (2003). VPC is an inclusive term referring to any two-word verb which consists of a verb and a particle in order to avoid the confusion brought about by definitions. Lam (2003:73-74) states that VPCs can be transitive or intransitive, transparent or figurative, and many of them are polysemous. He also points out that figurative VPCs are often referred to as PVs. Therefore, PVs are viewed as one subcategory of VPCs (Lam, 2003:75).

Besides the alternative names, the second problem is that the term 'phrasal verb' is also used to cover other types of MWV such as 'prepositional verb' or others (Dixon, 1982; Sroka, 1972). The inclusion of other MWVs is possibly a consequence of their complicated nature and the definitions given by the researchers. These inconsistencies lead us to consider the definition of PVs.

Traditionally, PVs are often defined as "idiomatic combinations of a verb and adverb, or a verb and preposition, or verb with both adverb and preposition", as Courtney (1983:1) puts it in the *Longman Dictionary of Phrasal Verbs*. Such a

definition is also applied by Cowie (1993:39), where PVs are “idiomatic combinations, whether of verb + adverb or verb + preposition”. Similarly, in *The Grammar Book* (Celce-Murcia & Larsen-Freeman, 1999:265), PVs are defined as “a verb followed by a particle variously described as a preposition, an adverb, or some combination of the two”. In the same vein, because the categorisation of PVs is complicated and entails multiple factors, Sinclair (2004a:162) also provides a broad definition of PVs in his *English Grammar*. He simply states that “the phrasal verbs consist of two or three words with adverbs or prepositions”. These definitions do not distinguish the adverbial or prepositional status of the particles. Perhaps the reason for the simplicity of these definitions is that they are designed to be understood by the learners who are the typical readers of grammar books and dictionaries. In spite of these definitions, however, not all scholars advocate the mixing of adverbs and prepositions in delimiting a single class of verbs.

In fact, other scholars have taken the opposite stance and made a sharp distinction between adverbial particles and prepositional particles. In the *Longman Grammar of Spoken and Written English*, Biber et al. (1999:403) give the definition of PVs as “multi-word units consisting of a verb followed by an adverbial particle”. They add that these adverbial particles have “core spatial or locative meanings” and

“are used with extended meanings”. Prepositional particles are thus excluded from the definition of PVs. Along the same lines, Claridge (2000:55) describes PVs as “combinations of a verb and a primary, invariable adverb, the latter including the heads of reduced prepositional phrases but excluding adpreps”, a term which Claridge (2000:49) adopts from Bolinger (1971) to refer to the particle which has ‘dual constituency’ in relation to the verb and the following noun phrase, as in “*He ran down it*” (=descend it). Therefore, an adprep is neither an adverb nor a preposition. Claridge only considers adverbs which imply ‘motion/ result’ as the acceptable particles in PVs. Giving more weight to the entirety of PVs, Darwin and Gray (1999:76-77) state that “a phrasal verb consists of a verb proper and a morphologically invariable particle that function together as a single unit both lexically and syntactically”. Note that they do not mean that the verb and its particle are inseparable when they say that they function as one unit. Schneider (2004:230) also advocates that “phrasal verbs are verb-particle combinations which are frequently semantically not transparent at all and strongly idiomatic, so the fusion of ‘two words’ to a new, complex lexical unit is practically complete”.

It is observed in the literature that two issues are controversial: (1) the separation of PVs and prepositional verbs (2) the separation of idiomatic PVs and free

combinations (Lindner, 1983, cited in Lam, 2003:76). Therefore, the next two sections will contribute to the discussion of the two problematic constructions, prepositional verbs and free combinations, and their relationship with PVs.

## **2.3 Other relevant MWVs**

### **2.3.1 Prepositional verbs and PVs**

The definition of prepositional verbs given by Quirk et al. (1985:1155) is: “A prepositional verb consists of a lexical verb followed by a preposition with which it is semantically and/or syntactically associated”. A prepositional verb such as the one in “*care for the parties*” is taken syntactically as a verb *care*, with a preposition *for*, followed by the complement of the preposition *the parties*. “*Care for*”, would not, like a PV, be regarded as a unified verb unit with a direct object “*the parties*” (Quirk et al., 1985:1156).

It is always difficult to differentiate ‘prepositional verbs’ and PVs (Claridge, 2000:47), especially transitive PVs, because they look very similar. An additional confusing point is that some prepositional verbs can have idiomatic meanings and act like one verb unit, for example instances such as *look after = tended*, *go into = investigated* (Quirk et al., 1985:1156), whose idiomatic nature is liable to be muddled with PVs.

Many linguists have endeavoured to distinguish between PVs and prepositional verbs. Quirk et al. (1985:1156) propose a test which consisted of moving the particle to the position after the noun phrase, because a true preposition in a preposition verb will not allow such movement (Rule 1). For example:

*She called on her friends. \*She called her friends on.* (prepositional verb)

*She switched on the light. She switch the light on.* (phrasal verb)

Quirk et al. (1985:1167) also give more rules to distinguish PVs and prepositional verbs, such as: (2) A pronoun is put before the particle in a PV but after the particle in a prepositional verb; (3) An adverb can be inserted between the verb and the particle in a prepositional verb but not in a PV; (4) The particle of a PV never occurs before a relative pronoun but is possible for a prepositional verb (e.g. *\*The man up whom they called. The man on whom they called.*); (5) The particle of a PV also never occurs before a *wh*-question, but the particle of a prepositional verb does (e.g. *\*Up which man did they call? On which man did they call?*); (6) The particle of a PV normally receives the stress. Similar tests like these, set up to distinguish PVs and prepositional verbs or to define PVs, can also be found in many other studies: details can be seen in Bolinger (1971), Fraser (1974, 1977), Darwin and Gray (1999), Claridge (2000), Sawyer (2000), Lam (2003) and Schneider (2004). Some of these

tests will be reviewed at length in Section 2.5.2.3.

Although these tests are able to separate PVs and prepositional verbs in most cases, there are exceptions and ambiguities. An example is *put up*, pointed out by Bolinger (1971:119). Compare:

*He put up a good fight.* (a show of resistance, a good argument)

\**He put a good fight up.*

In these two sentences, *put up* may be taken as a prepositional verb because of Rule 1 that a prepositional verb forbids reversal of the particle and the NP but a PV allows it (for more examples, see Lam, 2003:82). It is more reasonable to take it as a PV because it is idiomatic/opaque and the particle cannot be repeated as in: \**He put up a good fight, and up a good argument.* Quirk et al. (1985:1157) also acknowledge some special cases which worsen the confusion between PVs and prepositional verbs. An instance like *turn on* can be a PV (= *excite someone*) and prepositional verb (= *attack someone*). In such a case the meaning is changed, but in another case like *run over*, the meaning is similar, as in *The car ran him over.* (PV) *The car ran over him.* (prepositional verb), where *run over* can also act as both types of verb. These homographs exacerbate the difficulties in differentiating PVs from prepositional verbs.

Another extra problem in isolating prepositional verbs is the common construction ‘verb + preposition’ which has the identical form to a prepositional verb. Consider the example *He stayed at the corner*, in which the preposition *at* is part of the prepositional phrase *at the corner*, but is not associated with the verb *stayed*. Such kind of constructions should not be examined in the data of this thesis, because the preposition does not form a unit with the verb and associate with the verb directly. Quirk et al. (1985:1163-1164) provide some ways in which to isolate the common ‘verb + preposition’ constructions from prepositional verbs. The preposition of a prepositional verb can be fronted (e.g. *He called on her. On whom did he call?; He called before lunch. \*Before when did he call?*), and an adverb can be inserted between the verb and the preposition (e.g. *He called unexpectedly on her*), also a prepositional verb can be passive (e.g. *She was called on. \*Lunch was called before.*).

Although so far we have seen that some scholars such as Quirk et al. (1985) and Claridge (2000) advocate the separation of PVs and prepositional verbs, some other researchers have opted to fuse these two types of verb. Cornell (1985:279) includes both prepositional verbs and phrasal-prepositional verbs (see Section 2.3.3) in his discussion of PVs. Likewise, according to the *English Grammar* (Sinclair, 2004a), the function of the particles of phrasal verbs is to extend or change the meaning of a verb.

The particles can serve three functions: (1) as an adverb, e.g. *sit down* (2) as a preposition, e.g. *look after* (3) the verb can also be followed by both an adverb and a preposition, e.g. *look forward to*. These researchers agree on including prepositional verbs in PVs.

Some other scholars appear to have an inconsistency in their position. As mentioned earlier, Celce-Murcia and Larsen-Freeman (1999:265) consider PVs to have a particle which can behave like a preposition, an adverb or some combination of the two. Such a view seems to allow PVs to have a prepositional particle: they include *look into* (ibid.:265) and *come across* in their examples of PVs. Controversially, they suggested distinguishing PVs and prepositional verbs (ibid.: 268). The two aforementioned examples, *look into* and *come across* should be classified as prepositional verbs according to their rules, which are similar to those in Quirk et al. (1985:1167), and Celce-Murcia and Larsen-Freeman (1999:269). However, they are taken as phrasal verbs earlier in their book, as we have seen.

### **2.3.2 Free combinations and literal phrasal verbs**

Another type of MWV which is also easily confused with PVs is the ‘free combination’, which exhibits similar surface forms to PVs. The term ‘free combinations’ is proposed by Quirk et al. (1985:1152); they consider them to be the

combination of a lexical verb and an adverb, where both elements have ‘distinct meanings’. That is to say, the verb and the adverb each have their own meaning. These are exemplified as *He walked past (past the place)* and *I waded across (across the river)*. Most importantly, the adverb can be substituted by other adverbs, e.g. *He walked past/along the place*.

Researchers such as Dixon (1982) and Quirk et al. (1985) differentiate PVs from ‘free combinations’. Quirk et al. (1985:1152) list three methods to distinguish PVs from free combinations. First of all, the meanings of PVs cannot usually be predicted from the combination of the verb and the particle, while in free combinations they can be inferred from the verb and the adverb. Moreover, unlike PVs which function like a whole unit, both elements in free combinations, the verb and the adverb, have their own meanings. Either of them can be substituted by other lexical items, for instance, *put + down/outside/away; take/turn/bring + out*. It is also possible to insert an adverb such as *right* or *straight* between the adverb and the verb in free combinations, but this is unacceptable for PVs, e.g. *go right on, walk straight in*. Another syntactic characteristic is also suggested to differentiate PVs and free combinations: the possibility of positioning the adverb in the first place in a subject-verb inversion sentence for free combinations, e.g. *Out came the sun*.

Unfortunately, these methods are not without their problems. First, although the constituents in a PV cannot normally be replaced, in fact, there are some possible substitutions in an authentic PV such as *turn out the light*. Both the verb and the particle here can be replaced by other words (Quirk et al., 1985:1154):

*Let's switch it off.*

*Let's put it down.*

Second, there are also some ambiguous cases which cannot account for the last criterion satisfactorily. It is generally true that a free combination allows the particle to be fronted, but a rare case such as *\*Away they chattered* is not acceptable (Quirk et al., 1985:1153). Third, among the examples of 'free combinations' given by Quirk et al. (1985:1152-1153, 1162) such as *go on, drink up, walk in, come out, chatter away, bring in, take out*, etc., some instances e.g. *drink up, chatter away*, are 'semi-idiomatic constructions' (ibid: 1162). The boundary is not clear when these instances are concerned. Moreover, with the same form, some PVs can act transitively or intransitively in different meanings, for example, *give in = yield*, but *give something in = hand in*. This further complicates the judgment of PVs and free combinations.

Furthermore, in Quirk et al.'s (1985:1152) opinion, the most distinctive characteristic a free combination has that distinguishes it from a phrasal verb is that a PV is a whole unit with an unpredictable meaning. Such a distinction is not valid, because PVs are not always opaque. Some PVs are literal and transparent, which means that their meaning can be interpreted easily. For example, *sit down* is such a literal PV but not a free combination. Although, like a free combination, it is unidiomatic and denotes a direction of motion, the verb and the particle cannot be substituted freely in that given sense. The bond between *sit* and *down* is tighter than that between *put* and *down* in *Please put the cup down*. As these differentiation methods can fail, careful examination is required when attempting to distinguish PVs from free combinations.

Therefore researchers like Celce-Murcia and Larsen-Freeman (1999), Biber et al. (1999:403) and Lam (2003:76) consider free combinations and PVs as one group. Celce-Murcia and Larsen-Freeman (1999:267) do not state explicitly that they intended to combine the two constructions, but they regard the free combination *throw away (the ball)* as a phrasal verb example, which suggests that they were combining the two. Biber et al. (1999:403) warn that free combinations cannot practically be isolated, because fixedness is graded and not discrete. Lam (2003:76) groups *pull (the*

*curtain) down and blow the place up* together because he considered them to be “close enough”. He also observes that free combinations are usually not listed in the dictionaries of PVs because they are not regarded as real PVs, or because of the huge number of possible free combinations which makes it unnecessary to list them (Lam, 2003:80).

I agree with these researchers that free combinations and PVs are not separable. Researchers (Dixon (1982) and Quirk et al. (1985), as seen above) who advocate dividing these two groups are based on idiomaticity/opaqueness or wholeness/in-substitutability. As regards idiomaticity, Lindner (1983, cited in Lam, 2003:80) also argues that free combinations and PVs are just the two ends of the continuum of idiomaticity. It does not make sense to divide one family into two groups. Regarding in-substitutability, we have seen earlier that some exceptions break the rules. In consequence, it is better to discuss these two groups as one.

### **2.3.3 Phrasal-prepositional verbs**

One subtype of multi-word verbs which is usually included in PVs is the ‘phrasal-prepositional verb’. This construction comprises a verb, an adverb and a preposition (Quirk et al., 1985:1160), e.g. *look forward to, put up with, get away with*, etc. Phrasal-prepositional verbs are indistinguishable from PVs because they often

consist of a phrasal verb with a preposition, thus can be taken as the extension of two-word PVs. After reviewing other relevant types of MWV, we will turn to examine how PVs are treated in the dictionaries.

## **2.4 PVs in dictionaries**

One piece of evidence that PVs can be considered as a special group apart from other lexical items in English is the publication of dictionaries that focus exclusively on PVs, through which NNSs often learn PVs. Examples of phrasal verb dictionaries available at present include the *Oxford Phrasal Verbs Dictionary* (McIntosh, 2006), *Cambridge Phrasal Verbs Dictionary* (McCarthy & Walter, 2006) and *Collins CoBUILD Dictionary of Phrasal Verbs* (Sinclair, Hanks, & Moon, 2002). However, some problems were found in those dictionaries. First, most of them do not provide a clear definition of what a phrasal verb is. Second, the usages of the subcategories of MWVs are quite different.

As a result, the compilation of these dictionaries is not coherent to a certain extent, because their principles of selection are not clear. For example, two similar verbs *climb up* and *rise up*, both indicating actions toward the upward direction, were surveyed in the three mentioned dictionaries. The former is omitted by all the three dictionaries but the latter is included in two dictionaries. Inconsistency in choosing

PVs is not the only problem between dictionaries. The other problem is whether the same construction with the literal meaning should be counted as a PV. Take the example *wrap up*: it can be literal ('to cover something in paper, cloth...') or idiomatic ('to complete an activity'). Both senses are recorded in the dictionaries. If the literal combinations are also phrasal verbs, then why is *climb up* not regarded as a PV, given the fact that it is generally not included in dictionaries as seen. Moreover, the inclusion of prepositional verbs can be found in many dictionaries of phrasal verbs. Unlike researchers such as Quirk et al. (1985) and Claridge (2000), who have advocated the separation of PVs from prepositional verbs, the editors of dictionaries of PVs hold a looser view that allows all three categories of multi-word verbs in the family of 'Phrasal Verbs'.

In order to test whether the inconsistency only exists in traditional dictionaries and whether the classifications of PVs by researchers differs from those by dictionary compilers, I further examined five random examples: *walk past*, *go on* and *drink up* (classified by Quirk et al. (1985) as free combinations), and *pull away* and *yield up* (given by Claridge (2000) as literal phrasal verbs). I checked these five combinations in one traditional paper-based dictionary, the *Cambridge Phrasal Verbs Dictionary* (CPVD) (2006), and two online dictionaries: *Using English* and *Phrasal Verb Demon*.

The discrepancies are shown in Table 2.2, from which two phenomena are observed: the first is that ambiguous cases of PVs are always present, both in the paper-based and online dictionaries. Second, some of these verbs (*walk past*, *pull away*) seem to achieve more agreement than others (*go on*, *drink up*, *yield up*). It appears that the gaps between linguists and dictionary editors are obvious in cases such as the last three examples.

**Table 2.2: Survey of the inclusion of PVs in research and dictionaries (N=No, Y=Yes)**

	<b>Quirk</b>	<b>Claridge</b>	<b>CPVD</b>	<b>UE</b>	<b>PVD</b>
Walk past	N	--	N	N	N
Go on	N	--	Y	Y	Y
Drink up	N	--	Y	Y	Y
Pull away	--	Y	Y	Y	Y
Yield up	--	Y	Y	N	N

## **2.5 Elements, features and classification of phrasal verbs**

We have discussed the related constructions of MWVs that cause difficulties in identifying PVs. In this section I focus on the construction of the phrasal verb itself.

The elements of a PV will be discussed first, and then the syntactic and semantic features of PVs will be probed. The tests used to identify PVs in the literature will also be reviewed.

### 2.5.1 Elements

Theoretically, any verb can be the first constituent in a phrasal verb, but the possible candidates for particles form a rather closed group. Therefore, it is worthwhile spending some time considering the status of the particles. The possible particles are summarised from Quirk et al. (1985:1511) and Claridge (2000:46) in Table 2.3. The particles in bold were retrieved from the corpus data and added by Claridge, and those not in bold are identical in both reports by Quirk et al. (1985) and Claridge (2000).

**Table 2.3: Examples of particles**

	<b>Quirk et al. (1985)</b>	<b>Claridge (2000)</b>
Either adverbs or prepositions	About, above, across, after, along, around, by, down, in, off, on, out<AmE>, over, past, round, through, under, up	<b>Aboard</b> , about, above, across, after, along, around, <b>behind</b> , by, down, in, off, on, over, past, round, through, <b>to</b> , under, up
Adverbs only	Aback, ahead, apart, aside, astray, away, back, forward(s), home, in front, on top, out<BrE>, together	Aback, ahead, apart, <b>ashore</b> , aside, astray, <b>asunder</b> , away, back, <b>counter</b> , <b>forth</b> , forward(s), home, out, together

It is notable that Quirk et al. (1985:1162-1163) identify the groups of MWVs which have common meanings shared by the same particle. For example: they consider that *away* carries a meaning of ‘persistent action’, *up* suggests ‘completion’, *around*

implies ‘aimless behaviour’ and *out* means ‘endurance’. Sinclair (1991:68) also suggests grouping phrasal verbs by their particle in order to make “sense groupings”.

This idea is very insightful for teaching English to language learners.

## **2.5.2 Features and classification**

### ***2.5.2.1 Syntactic features: transitivity and separability***

Two classificatory approaches to distinguish PVs can be found in the literature: syntactic classification and semantic classification. At the syntactic level, phrasal verbs can be classified as transitive or intransitive, although some cases can be either transitive or intransitive, for example: *give in*, *blow up* (Celce-Murcia & Larsen-Freeman, 1999:266; Quirk et al., 1985:1152-1153). Besides transitivity, the other syntactic condition is whether the PV is separable. Examples of inseparable PVs include *get off* (‘descend from; leave’), *turn into* (‘become’), etc. Other PVs such as *cut off* (‘interrupt; sever; amputate’), *hand down* (‘deliver; pronounce formally; leave as an inheritance’) are separable.

Researchers such as Cowie (1993) and Hampe (1997) have attempted to elaborate these problems of the complex syntactic features of phrasal verbs. Traditional studies of PVs were conducted with a focus on syntactic complexity (Dehé, 2002; Dixon, 1982; Farrell, 2005; Fraser, 1977; Johansson, 1975; Sroka, 1972)

through introspection, while other scholars drew on corpus evidence (Gries, 2003; Potter, 2005). Other research has aimed to generate approaches to separating PVs from other verbs by running several syntactic rules (Darwin & Gray, 1999). The complication of the syntax has also meant that PVs are a tricky problem for computational linguists (Baldwin & Villavicencio, 2002; Berry-Rogghe, 1974; Li, Zhang, Niu, Jiang, & Srihari, 2003; Villavicencio, 2003) in search of an efficient approach to extracting PVs automatically.

#### ***2.5.2.2 Semantic features: idiomaticity of PVs***

At the semantic level, the most essential feature is idiomaticity. The term ‘idiomaticity’ is used rather incoherently in the literature. The concept ‘idiomaticity’ can be defined in a broad sense, which applies to studies of phraseology in the text, or in a narrow sense, which only accounts for specific language phenomena, such as phrasal verbs.

Summarised from a number of previous studies, the term ‘idiomaticity’ incorporates several key points:

1. Idiomaticity (or non-compositionality) is a feature of phraseological units, which states that the meaning of the whole unit cannot be deduced by combining every single lexical item. In other words, the meaning of an idiomatic PV is opaque (Waibel, 2007).

2. Idiomaticity is a cline with increasing degrees of semantic opacity and structural stability (Cowie, 1998:213).
3. Idiomaticity is both (1) nativelike selection of expression, and (2) that which one has to know over and above rules and words (Warren, 2005:35).

By and large, idiomaticity is taken to be the characteristic of non-randomness or phraseology in the text, but it is commonly mixed with the study of idioms (Prodromou, 2003; Warren, 2005). It is also often regarded as an indicator of language users' proficiency or nativeness for L2 learners. In its narrow sense in relation to PVs, concepts such as non-compositionality (Waibel, 2007:5), non-literalness (Waibel, 2007:15), semantic complexity or opacity (Armstrong, 2004:215), etc., have all been suggested by researchers. Most of these terms are synonymous and interchangeable. To put it simply, idiomaticity refers to the fact that the meaning of the PV cannot be inferred by combining its individual constituents.

Idiomaticity is essential for defining a PV for some researchers. For example, a PV is defined as "a verb + particle combination that functions as a single verb, both parts giving up meaning in order to form a new lexical item" (Darwin & Gray, 1999:65). Note that in this definition, the meanings of the verb proper and the particle

are diluted, and a new meaning is created, either deriving from the original item or generating a different new one. Therefore, the meaning is opaque and cannot be inferred easily from the combination of the verb and the particle. Of course, this refers only to idiomatic PVs. If we take idiomaticity as a continuum, then VPCs with no or less idiomaticity/opaqueness can also be PVs. They are often termed literal/transparent and figurative PVs in the research.

Studies have been conducted in an attempt to pin down the semantics of PVs (Armstrong, 2004; Consigny, 2001), or to cover both meaning and structural heterogeneity (Televnaja, 2004). By virtue of these complexities of PVs, a number of studies have contributed to the categorisation of PVs. Previous classification approaches of PVs vary according to the purpose of the research. One semantic classification of PVs is to divide them according to which parts contribute to the meaning of a particular PV (the verb, the particle or the whole unit); thus PVs are grouped into ‘verb + adverbial particle’, ‘verb + *aktionsart* particle’, and ‘non-compositional, idiomatic PVs’ (Konig, 1973:90, cited in Claridge, 2000:55). A similar view is adopted by Armstrong (2004:222), who also divides PVs according to their compositionality. He terms his three types ‘directional PVs’, ‘aspectual PVs’ and ‘idiomatic PVs’. Some other linguists have taken a different view. For example, Quirk

et al. (1985:1162) divide PVs into three categories: ‘free, non-idiomatic constructions’ (*take out, walk up*), ‘semi-idiomatic constructions’ (*beaver away, finish up*) and ‘highly idiomatic constructions’ (*bring up, turn up*), according to the degrees of possibility of substituting one element, either the verb or the particle. Laufer and Eliasson (1993) also provide a classification scheme for PVs based on their semantic properties. They categorised three types of PV: (1) semantically transparent, e.g. *come out* (2) semantically semi-transparent, e.g. *let down* (3) semantically opaque/idiomatic or figurative, e.g. *put off*. Along similar lines, Celce-Murcia and Larsen-Freeman (1999:274) group PVs into ‘literal’, ‘aspectual’ and ‘idiomatic’, which are parallel to the generally recognised ‘transparent PVs’, ‘semi-transparent PVs’ and ‘opaque PVs’. Liao and Fukuya (2004:196-197) divide PVs into (1) literal: *go out, take away, come in, get up, go away* (2) figurative: *turn up, let down, show up, go off, hold on, put out, make up, give in, turn down, show off, run into* (3) completive: *cut off, burn down* (for a similar classification, see also Dagut and Laufer, 1985:74). Such a classification emphasises semantic transparency more than compositionality. These classifications are summarised in Table 2.4.

The parameters which have been used to select PVs in the sub-categories include semantic nature and commutability; these two criteria are pointed out by Howarth

(1996). Quirk et al. (1985) deal with classification by measuring the substitutability of the verb-particle constructions. Other researchers have tried to divide them by their semantic features: idiomaticity and the addition of new meaning. They have decided the classification based on whether the collective meaning can be obtained by combining the individual elements, and whether the PV carries metaphorical meaning, or whether the particle suggests completeness (see above, completive PVs in Liao & Fukuya, 2004). It must be noted that when we explain PVs by idiomaticity, the classification can be complicated by the additional meanings carried by the verb or the particle. In some cases, the meaning is added by the particle, as remarked by Side (1990:146): “in all phrasal verbs the particle carries some meaning. In many, it carries most of the meaning”. Or as argued by Consigny (2001:239), both the verb and the particle contribute to the meaning but “neither has any kind of dominance”. In any case, the extra meaning will have an impact on the classification.

**Table 2.4: Classification of PVs in previous studies**

Author	Categories
Dagut and Laufer (1985:74)	<i>literal PVs</i> : meaning is combined by the verb and the particle <i>figurative PVs</i> : metaphorical shift of meaning <i>completive PVs</i> : the particle indicates the result of the action
Laufer and Eliasson (1993)	<i>semantically transparent</i> : meaning derives from combining the two parts <i>semitransparent</i> : meaning becomes transparent in context <i>semantically opaque/figurative</i> : meaning is lexicalised
Liao and Fukuya (2004)	Definitions are not given in the study: <i>literal</i> <i>figurative</i> <i>completive</i>
Celce-Mercia and Larsen-Freeman (1999)	<i>literal</i> <i>aspectual</i> <i>idiomatic</i>
Quirk et al. (1985)	Three categories are identified by the substitution of one element, either the verb or the particle: <i>Free, non-idiomatic constructions</i> <i>Semi-idiomatic constructions</i> <i>Highly idiomatic constructions</i>

There are two basic problems in classifying PVs by degree of idiomaticity. The first involves determining the degree of idiomaticity of a phrasal verb. For this reason, the issue of idiomaticity has been largely put aside in previous studies. Berry-Rogghe (1974) attempts to measure idiomaticity: a statistical approach is used to define the idiomaticity of PVs. A VPC is taken as idiomatic if it has collocates different from

those which collocate with the particle alone. This approach provides a way to capture idiomaticity, but generally speaking, the degrees of idiomaticity are relative and controversial in many cases. The problem of defining PVs by their degrees of idiomaticity lies in the fact that no clear-cut boundary can be drawn between the categories of PVs. The degree of idiomaticity is usually a matter of relativity. The scale of idiomaticity is a continuum of gradience without clear-cut boundaries.

Alongside the problem of determining how idiomatic a PV is, the confusion of labels such as 'literal' and 'figurative', etc. also result in more complexity. Waibel (2007:18) points out that the definitions of 'transparent', 'idiomatic', 'opaque', 'figurative' and 'literal' need more detailed differentiation. She produces a table to compare these notions and to see the extent to which they are similar to each other. In conclusion, she takes 'idiomatic' and 'opaque' as being similar terms and finds that 'literal' contrasts with 'figurative', 'opaque' and 'idiomatic', while 'transparent' runs counter to 'idiomatic' and 'opaque'. Therefore, idiomaticity and transparency are regarded as taking oppositional positions. The definitions used in her research are listed below. Although she does not provide explicit definitions for 'literal' and 'figurative', her ideas about these two terms can be found from the examples she cites.

- idiomatic/opaque: the meaning cannot be derived from the combined meaning of its parts
- transparent: the meaning is not concealed
- literal: the meaning of the particle involves the directional, spatial or locative senses
- figurative: the meaning does not involve the actual, physical aspect (Waibel, 2007:17-18)

The terms can also be easily confused with the general, non-technical uses of the words involved. For example, ‘idiomatic PVs’ are not equal to ‘idioms’. The elements of an idiom are irreplaceable, but an idiomatic PV may take many different collocates. Therefore, I will use ‘idiomatic’ to mean that a PV is ‘non-compositional and semantically opaque’. Furthermore, besides the conflicts which have been noted, Waibel (2007) points out that the denotation of ‘figurative’ is also problematic. It can refer to a traditional figurative use where a literal PV, an action verb with a directional particle, is applied to an abstract proposition (e.g. *bring students back to school*). Alternatively, it can also suggest that a special meaning is added to the PV. For example, many extended senses of the particle *up* have been identified by cognitive linguists, such as ‘reaching a goal/end/limit’, ‘positive evaluation’, ‘higher in rank’,

‘more visible’, etc. and these can be said to make the PV more figurative (Neagu, 2007:133). Therefore, labelling a PV as figurative will make the situation more complicated. So the two terms ‘literal’ and ‘idiomatic’ will be adopted as interchangeable with ‘semantically transparent’ and ‘semantically opaque’, but ‘figurative’ will be regarded as another level to account for PVs, therefore the rest of the PVs apart from ‘literal’ and ‘idiomatic’ ones will be referred to as ‘semi-transparent PVs’ in this thesis.

A special group of PV also recognised by linguists in the literature is the ‘completive or aspectual PV’. Instances such as *drink up*, *cut up*, *eat up* all imply the status of ‘completing’, which is added by the particle *up*. These PVs are different from other categories, because whereas they share some properties, e.g. the meaning can be inferred (as transparent PVs) and they have a more fixed relation between the verb and the particle (as literal PVs), some idiomaticity is involved, because of the completeness suggested by the particle. Another important point is raised by Bolinger (1971:16): “the literal uses lie at the core, and figurative ones surround them at varying distances”. The same perspective will also be adopted in this thesis.

The issue of idiomaticity raises the question of including/excluding PVs with fewer or no degrees of idiomaticity. The issue concerning semantic transparent PVs

such as free combinations has been raised in Section 2.3.2. This issue will be addressed further at this point. Some scholars such as McArthur (1992) and Claridge (2000) agree that literal combinations should be included in the world of PVs. Claridge (2000:47) calls these ‘completely literal types’, in contrast to figurative, idiomatic combinations. She considers that literal types should be included for two reasons. In her view, idiomaticity derives from the core of these literal combinations. That is, a figurative sense, say *wrap up a meeting*, usually derives from the literal sense, *wrap up gifts*. Furthermore, she believes that there is a pragmatic cline where it is difficult to draw a clear dividing line between the literal and idiomatic ends.

Apart from idiomaticity, another issue around semantic features which calls for our attention is the polysemy of PVs, or to put it in another way, the problem of homographs. Consider the examples:

- Blanche *put down* her cup. [BNC CDY 1719]
- Stealthily again she *put down* the phone. [BNC AE0 2371]
- The three-year-old alsatian, called Sam, was at an animal shelter in Gosforth, Newcastle upon Tyne, yesterday waiting to be *put down*. [BNC CBF 574]

Some researchers take the three examples as one PV which has three meanings, thus is polysemous, and others take them as three different PVs or homographs. For the former position, one PV can behave as a literal, semi-transparent or idiomatic verb-particle combination, as illustrated by these examples. In the case of *put down*, we can see the gradient cline from a transparent meaning to semi-transparent and to the more opaque meaning: *put down a cup* ‘put something onto a surface’, *put down the phone* ‘put the phone back into its usual position’, *put down an animal* ‘kill the animal’. For the latter position, these are three individual PVs, each with different degrees of idiomaticity. No matter which position is held, this makes the study of PVs more complex. The semantic features of idiomaticity and polysemy/homograph confound the analysis of PVs.

From the literature concerning the idiomaticity of PVs summarised above, a critical problem can be noticed. Some researchers seem to conflate the notions of idiomaticity and commutability. For example, Quirk et al. (1985) grouped PVs by their substitutability but named them by their idiomaticity. On the other hand, others distinguished the two notions clearly. An example of such a distinction is Howarth (1996). This inconsistency has raised considerable confusion in the study of the collocation of PVs and will be addressed at length in Section 6.4.1.

### 2.5.2.3 *Tests to identify PVs*

Traditional studies of PVs have been conducted to define the membership of a phrasal verb using the approach of running a number of tests. Bolinger (1971) aims to prove that a verb + particle combination is indeed a phrasal verb, and proposes methods such as replacement, formation of passives, formation of action nominals, object movement, pronoun placement, adverbial insertion, stress, definite noun phrases and listing. Other researchers have also proposed a number of ways to identify PVs either syntactically or semantically. Lam (2003:81-94) makes a clear summary of the tests, which he categorised by their relation to the adverbial property of particles, the unity between verb and particle and the unity of prepositional phrase. The tests and examples are presented below:

- **adverbial property of particles**

1. NP-insertion test: The object NP can be inserted between the particle and the verb of a PV. (e.g. *He will look the client over.* \**He will look the fence over.*)
2. Particle-stress test: The verb of a PV receives the stress. (e.g. *He RAN up the hill.* \**He ran UP the hill.*)

- **unity between verb and particle**

1. Replacement test: A PV can be replaced by one word (usually Latinate).  
(e.g. *give in=yield, count out=exclude*)
  2. Passivisation test: The transitive PVs can be passivised. (e.g. *The place was blown up. \*Some letters were came across.*)
  3. Action nominal test: A PV can be nominalised. (e.g. *His carrying out of the work was surprising. \*His running up of the hill was stupid.*)
  4. Verb-insertion test: A PV does not take a verb between the verb and the particle. (e.g. *\*I messed and fouled up on my test.*)
  5. Adverb-insertion test: Non-PVs can have an adverb in between, but not PVs. (e.g. *\*She turned slowly up. She turned slowly away.*)
  6. Intonation unit test: No pause is allowed between the verb and the particle in a PV. (e.g. *\*I passed/out in the doctor's office.*)
- **unity of prepositional phrase**
    1. Verb-gapping test: The second head verb cannot be omitted in a PV. (e.g. *\*I looked up your name, up her name, and up his name.*)
    2. Where question test: (e.g. *Where did you look? \*Up the address. Where did he run? Up the alley.*)
    3. Particle fronting test: The particle of a PV cannot be fronted. (e.g. *Up he*

*made a story. Up the tree he went.)*

4. NP-ellipsis test: The object of a PV cannot be omitted. (e.g. *We turned off (the road).* \* *We turn off (the light).*)

However these methods are only valid to a certain extent. As exceptions can always be found among them, these methods are better taken as principles rather than hard-and-fast rules. The crucial problem caused by the complexity of these methods is the lack of consistent criteria with which to interpret data, which results in different results from different studies. To solve this problem, an alternative approach is proposed by Darwin and Gray (1999:65). They suggest that “instead of requiring verb + particle combinations to demonstrate specific features in order to be identified as phrasal verbs, the new approach calls for researchers and teachers to consider all verb + particle combinations to be potential phrasal verbs until they can be proven otherwise”. Such a proposal provides new insight into the way in which our perspectives of phrasal verbs should be reshaped. Their new tests to single out non-phrasal verbs are: (1) particle repetition, (2) where questions, (3) fronting, (4) verb insertion, (5) adverb insertion, (6) stress, (7) intonation units. However, because we are dealing with written texts, only the first five tests are applicable. These five

tests listed below will be employed in this thesis to discard the non-targets where necessary.

- **Particle repetition**

- PV: *\*I looked up, up, up your name.*
- Non-PV: *I looked up one aisle, then up the next.*

- **Where questions**

- PV: *I looked up the address./ Where did you look?/ \*Up the address.*
- Non-PV: *He ran up the alley./ Where?/ Up the alley.*

- **Fronting**

- PV: *He made up a story./ \*Up a story he made.*
- Non-PV: *He went up the tree./ Up the tree he went.*

- **Verb insertion**

- PV: *I really messed up on my test./ \*I really messed and fouled up on my test.*
- Non-PV: *He pulled on the lever, but it was stuck./ He pulled and jerked on the lever, but it was stuck.*

- **Adverb insertion** (NB: two adverbs must be used and they both have to

be -ly adverbs.)

- PV: *\*The mine caved quickly and forcefully in.*
- Non-PV: *They crept slowly and silently down the hall.*

## 2.6 Summary

In this chapter, we have looked at the subcategories of multi-word verbs. Phrasal verbs are one subcategory of MWVs and the identification of PVs is problematised when other similar constructions such as prepositional verbs and free combinations are concerned. Prepositional verbs can be separated from PVs by a number of tests, although a few cases which violate the tests can be found. Problems such as inconsistency in the studies also suggest the difficulties that may be encountered. Likewise, the identification of free combinations can be achieved by running tests, but again, these tests are not watertight. In addition to these problems, we can also see the contradictions within single studies. For example, scholars such as Quirk et al. have paid attention to classifying or delimiting PVs and other MWVs. They exemplify *switch on* as a PV in contrast to the prepositional verb *call on* (Quirk et al., 1985:1157). However, *switch on* is not regarded as a PV but a free combination, according to their definition that the verb and the particle can be substituted by other words like *turn* and *off* (Quirk et al., 1985:1152). Therefore, in conclusion, there is no point in attempting to

isolate free combinations from PVs.

The argument about including or excluding prepositional verbs and free combinations can be observed by the different stances taken by previous researchers. Lam (2003:77-78) creates a table to divide these stances by whether the prepositional verbs are kept and whether only the idiomatic ones are maintained. Four types of division can be generated according to these two conditions. The four divisions can be represented by the presence (+) or absence (-) of prepositional verbs (P) and free combinations (F) as the following four groups: +P, +F; +P, -F; -P, +F; -P, -F. Each division has its own advocates and is named differently (see Lam, 2003:77-78). The researchers opt for the one which meets their research needs.

In this chapter, the semantic and syntactic features of PVs have been discussed and classifications based on these features have also been reviewed. The problems regarding the terms and the sub-classifications of PVs have been revealed, as have the problems of idiomaticity and polysemy/homographs. All of this adds complexity to the analysis of PVs. Finally, the tests used to identify PVs have been introduced, and these tests will be applied in our analysis process where necessary.

A table of the other types of MWVs which are easily confused with PVs was created (see Table 2.5 next page), showing their semantic and syntactic features. It is

hoped this table will help to clarify any confusion. Chapter 5 will discuss in detail how the target PVs in this thesis are to be identified. In the next chapter, I will review the literature relating to learner language to bring out descriptions of LL characteristics, the contributions of corpora to analysing LL, and the teaching/learning of PVs.

**Table 2.5: Properties of different VPCs**

Form	Term	Idiomacity	Transitivity	Separable	Single-unit	Example
V+ prep.	Common V+prep.		transitive	N	N	<i>stay at, walk on</i>
	Prepositional verbs	non-idiomatic	transitive	N	Y	<i>Look at</i>
		idiomatic	transitive	N	Y	<i>come across</i>
V+ adverb	PVs	Non-idiomatic (Free combination)	intransitive	N	N	<i>go out</i> <i>fell down</i>
			transitive		N	<i>walk across</i> <i>climb up</i>
		Figurative	intransitive	N	Y	<i>beaver away</i>
			transitive	Y	Y	<i>drink up</i>
	idiomatic	intransitive	N	Y	<i>buckle down</i>	
		transitive	N	Y	<i>come by</i>	
			Y	Y	<i>bring up</i>	



# **Chapter3: LEARNER LANGUAGE, CORPORA AND PHRASAL VERBS**

## **3.1 Introduction**

This chapter has two-fold aims: giving an introduction to what learner language is like, how corpora have contributed to examine learner language and second/foreign language pedagogy, as well as reviewing the findings of PVs in learner languages. It first provides a brief review of the general characteristics of the learner language, and then focuses on the relationship between phraseology and L2 acquisition, including studies of learners' collocational knowledge and the approaches which have been employed to describe learner language, particularly corpus studies. The application of corpora to teaching will be touched upon as well. The second half of this chapter will contribute to the related issues of PVs in learner language, including problems experienced by learners, and the roles of PVs in previous studies.

## 3.2 Learner language

### 3.2.1 LL characteristics

Learner languages (hereafter LL) are often described as ‘unnatural’ in comparison to native English. The features of the ‘un-naturalness’ include speech-like writing, use of a smaller range of vocabulary items, or use of less specific vocabulary items, a tendency towards unidiomatic combinations, and fossilised errors (Guo, 2006). Two of these features, the use of general vocabulary and the lack of formulaic expressions, are of particular relevance here. General, common, vague and high-frequency words are reported to be favoured by learners, and at the same time only a limited range of vocabulary items are used (Granger & Rayson, 1998; Ringbom, 1998). This implies that the LL lacks specificity and elaboration of word meanings. Learners may not be able to make vocabulary selections as precisely as NSs; they choose general and all-purpose words instead to avoid errors. For example, in a study of amplifiers, Granger (1998a:151) concludes that learners tend to use ‘all-purpose’ amplifiers such as *very* instead of others which end in *-ly*. The other feature related to this present work is prefabs or formulaic sequences. It has been demonstrated that learners tend to overuse fewer formulaic sequences, which are not the same items used by NSs (De Cock et al., 1998; Ringbom, 1998). Idiomaticity is generally considered as an indicator

of nativelike-ness, characterised by collocations, sentence stems, the use of phrasal verbs and so on (Yorio, 1989:68). The extent of idiomaticity is determined by the amount of phraseology shown in the LL. It will be interesting to understand whether learners use PVs in the same way as NSs do in terms of their selection, number and idiomaticity. The roles and functions of phraseology are summarised in Chapter 4, with a focus on collocation studies.

### **3.2.2 Learners' lexical knowledge**

The second or foreign language learners' competence in using phraseology involves the quality of their lexical knowledge (Liu & Shaw, 2001:171). In studies of L2 vocabulary acquisition, the essential factors which enable one to use a word are analysed by Nation (2001:27). Knowing a word means knowing its form, meaning and use. The meaning part contains the sub-factors: form and meaning, concept and referents, and associations. The use part comprises grammatical functions, collocations and constraints on use (register, frequency, etc.). These factors are referred to as lexical knowledge. Read (2004:211ff) also summarises three constructs of vocabulary knowledge, including precision of meaning, comprehensive word knowledge and network knowledge. Note that the concept of a 'word' has been newly defined by Firthian scholars such as Sinclair (2004b), who proposes the idea of 'units of meaning'

(cf. Section 4.3.1). This new idea does not undermine the construct of lexical knowledge, but helps to complement it.

Among the subcategories of lexical knowledge, collocational knowledge is crucial to learners. Learners need to know how words are combined or collocated. Competence in using adequate collocations is believed to enable learners to achieve nativelikeness (Lesniewska, 2006:96). Knowing the ‘collocability’ is regarded as an essential part of learners’ lexical competence, and a lack of this competence can result in “a serious loss of precision” (Howarth, 1998:162). As reported in Waller’s (1993) research, collocational errors are rarely present in NSs’ writings but are prevalent in those of NNSs. Learners’ collocational knowledge has been found to develop with their proficiency levels, and lexical collocations are more difficult to acquire than grammatical collocations (Gitsaki, 1996). However, the maturation of their collocational knowledge does not keep pace with the growth of their knowledge about individual lexical items (Bahns & Eldaw, 1993).

Liu and Shaw (2001:188) examine learners’ lexical knowledge of the word *make*, and find that learners do not use the same grammatical and semantic distribution as NSs do, so they call for teaching the full word potential in depth to the students. They not only advocate making comprehensive studies of one word, but also suggest an

integration of lexis and syntax, similar to ideas proposed in pattern grammar (Hunston & Francis, 1999).

### **3.2.3 Development of learner language studies**

The characteristics of learner language can be captured by various approaches. The most noticeable feature is their errors, which are often investigated through error analysis (James, 1998). Error analysis has perhaps had its heyday, but it also received severe criticisms. With the blossoming of corpus linguistics, error analysis may find a new way to grow, but the traditional method of analysing learners' errors has become insufficient. Another dominant classical approach is contrastive analysis, which compares and contrasts at least two languages. The idea of contrastive analysis has been applied by Granger to LL studies, and termed 'Contrastive Interlanguage Analysis' (CIA) (Granger, 1996:43), which makes a comparison between the original and target languages and comparison between their translation equivalents. Later, Granger modifies the CIA model into a comparison of native language with interlanguage or different languages, i.e. (1) NL vs. IL and (2) IL vs. IL (Granger, 1998a:12). For the former (NL vs. IL), overuses and underuses are the primary means to determine the interlanguage differences. For the latter (IL vs. IL), different varieties of learner corpora are dealt with, and the transfer from their mother tongues examined.

Furthermore, Granger suggests integrating the CA and CIA to form a comprehensive account to increase the validity of learner language research.

CIA has been adopted widely in learner corpora studies, based on a ‘Computer Learner Corpus’ (CLC) approach, to use Granger’s term (Granger, 1998a:6). She gives a summary of the basic features, and current analysis approaches of CLC (Granger, 2004). In her opinion, CLC distinguishes itself from other data collection types in SLA; it has advantages in size, variability, and automation (Granger, 2004:124ff). She also contends that the methodological framework at the heart of CLC rests mainly on CIA (Contrastive Interlanguage Analysis) and CEA (Computer-aided Error Analysis). CLC research can be classified according to its research design. Granger (1998a:15) notes that CLC research can be classified into ‘hypothesis-based’ and ‘hypothesis-finding’. She concludes that hypothesis-finding is more powerful to “gain totally new insights into learner language” (Granger, 1998a:16). One of the study types which can benefit from the ‘hypothesis-finding’ design is research on formulaic sequences. Granger and other scholars have tested whether learner languages are composed of ‘individual bricks’ or ‘prefabricated sections’ (De Cock et al., 1998:67), and investigated vague language that occurred as some phraseological combinations (De Cock et al., 1998:74-79). Their results conclude that learners use more prefabs than native speakers,

with different frequencies and functions. In a more recent article (Granger, 2005b), Granger pinpoints two mainstreams of research on phraseology. The first trend is the interest in distinctions between less fixed multi-word units and free combinations. The second trend concentrates on typical features, such as non-compositionality and fixedness of the formulaic sequences. In order to complement the lack of a broader overview on the phraseological phenomena, she ends up with a suggestion of incorporating the statistical approach with fine-grained linguistic analysis as filters to yield targets worthy of further investigation. This thesis thus follows this suggestion.

### *3.2.3.1 Corpus approaches to describing LL*

With the advance of technology, corpora have been applied to inform the theories and practice of second language acquisition. As such, in recent academic history, CLC studies have been fruitful in describing learner language. Here I will review a few significant learner corpora and a number of CLC studies on different aspects of learner language.

Pravec (2002) surveyed the background information of several learner corpora. Because of the space limitations, I will focus only on corpora of written texts. To name some of the present-day learner corpora, the International Corpus of Learner English

(ICLE), the Cambridge Learner Corpus (CLC) and the Longman Learners' Corpus (LLC), will be described below.

The ICLE is a project created by the Centre for English Corpus Linguistics (CECL), at Université Catholique de Louvain. The leading researchers include Sylviane Granger, Fanny Meunier, Estelle Dagneaux, Magali Paquot and Sylvie De Cock. It is an international project which collaborates with other researchers in many countries, such as China, Germany, etc., containing over fourteen varieties of learner languages. A comparable reference corpus, LOCNESS, containing both British and American English, was built as well. All the corpora of ICLE were compiled in the same format and designed using the same rules, in order to ensure comparability. Many studies have been conducted examining the data from ICLE. For instance, the directing researchers mentioned above have produced papers on many aspects of learner language (De Cock, 2000, 2001; Granger, 2005a; Meunier & Granger, 2008). Also, other researchers such as Kaszubski (2000) have tackled the phraseological issues found in a sub-corpus of ICLE. In addition, two projects, the Longitudinal Database of Learner English (LONGDALE) and the Varieties of English for Specific Purposes database (VESPA), both derived from ICLE, were launched in 2008. The LONGDALE project collects longitudinal learner data and the VESPA project deals with learner English for specific

or academic purposes. Besides these projects, the CECL is also directing a project on phraseology and discourse. Other learner corpora include the Cambridge Learner Corpus (CLC) and the Longman Learners' Corpus (LLC). The CLC is a collection of Cambridge ESOL exams by the Cambridge University Press. Also compiled on a commercial basis, the LLC comprises 10 million words and is used mainly to inform the content of textbooks. CLC and LLC are not publicly available whereas ICLE is, thus ICLE is more advantageous to researchers.

Whereas learner corpora can contribute substantially to the understanding of LL, they have some limitations. Learner corpora are deficient in providing information on learners' receptive ability, motivation and reaction to certain teaching methods; in addition, they are particularly criticised for their inability to discover what does not exist in the LL (Nesselhauf, 2004:131-132).

Despite these deficiencies, learner corpora can provide much evidence in describing languages. Studies based on learner corpora have probed the special characteristics of LL at many levels. Some researchers have examined grammar; some are interested in lexis and phraseology; others are attracted by discourse and stylistics issues. Granger and Arts (1998), for example, explore tag sequences in LL. In terms of lexis, Ringbom (1998) looks into vocabulary and reports that learner languages are more limited in

lexical choices; Granger and Rayson (1998) also analyse learners' patterns of grammatical categories. At the stylistics and discourse levels, learner language is found to lack lexical variation and have a tendency towards overstatement and wordiness (Lorenz, 1998:64). In the area of phraseology, De Cock et al. (1998) observe prefabs such as two-word combinations and vagueness expressions. All of these studies have attempted to approach LL using CLC techniques.

### **3.3 Corpora and teaching/learning**

#### **3.3.1 Corpora and teaching**

The CLC studies are bound to have a great influence on language pedagogy, including curriculum or syllabus design, and language teaching (Keck, 2004). They have not only changed the way a language is described but also worked as the resources to generate pedagogical materials (Hunston, 2002:137). How corpora can help to describe languages has been reviewed above; this section will discuss how corpora benefit language teaching and their limitations.

One teaching method which relies on corpora and has been applied practically is data-driven learning (DDL), developed by Johns (1991). Following the steps of observation, classification and generalisation, the teacher shows concordance lines to the students and leads them to discover the answers from the context (for details, see

Hunston, 2002:170ff). Either through planned tasks or free discovery, the students will be motivated to learn and obtain a clear picture of a language feature, which can be maintained in the long term. This process of learning is termed 'learning as research' by Johns (1991), but modified to 'learning as discovery' by Bernadini (2004:23) to stress that learners can be guided by their own interests. Although corpora are a powerful tool for learning, we must be aware that the pitfalls of applying the DDL technique in the classroom are that it may be time-consuming and not pay off (Hunston, 2002:178).

DDL usually makes use of corpora, so learner corpora may also constitute resources for learning. Nesselhauf (2004:139) states that learner corpora can inform instructors about what and how to teach, appropriate sequences to introduce linguistic features, and probable mistakes. However, she also warns that unlike native corpora, which show adequate information, learner corpora may provide 'negative evidence' to the students, thus efforts must be put to direct them to notice the correct answers (Nesselhauf, 2004:140).

Besides offering innovations in teaching methodologies, corpora also shed light on pedagogic materials. The design of syllabuses is affected by corpora as well. In order to reflect what is authentically used in native English, Mindt (2000) proposes an 'empirical grammar' to illustrate a more effective way for learning the English verb

system. In addition to the syntactic aspect, corpus research has also brought forth the birth of the 'lexical syllabus'. Summarised in Hunston (2002:189), the lexical syllabus was first introduced by Sinclair and Renouf (1988), and was developed by Willis (1990). It is advocated that "(1) the commonest words in the language; (2) the central patterns of usage; (3) the combinations which they usually form" are concerned in a lexical syllabus (Sinclair & Renouf, 1988:148). The idea is to teach common words of high frequency and various usages, in order to enable learners to familiarise themselves with the sophistication of language with a widely-used word. These studies suggest analysing the aspects (grammar, lexis, etc.) of a word as 'patterns', which can be easily absorbed and intuitively applied by learners. Lewis (1993, 1997) puts forward the 'lexical approach', which can be practised in the classroom. The core of these activities is the Observe-Hypothesis-Experiment cyclical paradigm, which replaced the traditional Present-Practise-Produce paradigm (Lewis, 1993:6). Such a lexical approach advocates that "correctly identified lexical phrases can be presented to L2 learners in identifiable contexts, mastered as learned wholes..." (Lewis, 1993:96) and also emphasises the importance of 'idiomaticity' (Lewis, 1993:98) and 'contextualisation' (Lewis, 1993:103). The lexical approach seems to be promising in increasing learners' fluency, accuracy and ease of learning; however, Granger (2011:6)

is conservative about that such an approach providing sets of phrases to learners has limited 'generative power', which will result in risks of overloading the current teaching environment.

One of the problems of applying corpora to teaching is frequency. Corpus linguists agree that higher frequency denotes higher probability of usefulness. This is generally true, but some infrequent items may also be useful to learners. Cook (1998) notices that some expressions are rare but salient; Hunston (2002:194-195) further points out some infrequent items are important because they have 'cultural value'. These infrequent but significant words deserve more attention in pedagogic/reference materials.

For comparability and convenience, this thesis draws its data from academic corpora. When the corpora were selected, not many Chinese learner corpora were available. The use of academic corpora for investigating PVs is likely to result in a smaller amount of data because formal and written texts tend to contain fewer PVs, and certain PVs are more likely to appear in academic texts (cf. Section 9.3.1). If spoken and informal corpora had been selected, much greater frequencies and more literal uses of PVs could have been found. We will see later in Chapter 9 that factors such as corpora topics/genres/register have an influence on the uses of PVs. Therefore, choosing non-academic corpora would surely lead to different results.

### **3.3.2 Phraseology and teaching**

Drawing attention to teaching multi-word expressions is certainly the revolution that corpora bring to language education. This is reflected in studies which concern the implications of collocation and teaching, for example, Kaszubski (2000) and Nesselhauf (2003). For teaching multi-word units, it has been suggested by Wible (2008) that digital environments can benefit learners, especially in learning phraseologies, because digital resources are dynamic, distributed and active. Unfortunately, so far the key issues related to learning MWUs such as the frequencies of encounters and the kind of exposure that learners gain have not been clarified (Coxhead, 2008:155), and some problems remain which might pose challenges to teachers.

Ellis (2008:7-8) anticipates several potential challenges that the instructors may confront when teaching phraseology to learners. The first is that phraseology is acquired implicitly in a natural environment, but learners memorise formulaic language as explicit and declarative knowledge. Secondly, learning a new language may require learners to re-construe their world like NSs, but language transfer often hinders this progress. The final problem resides in how to learn the prototypical meanings of the

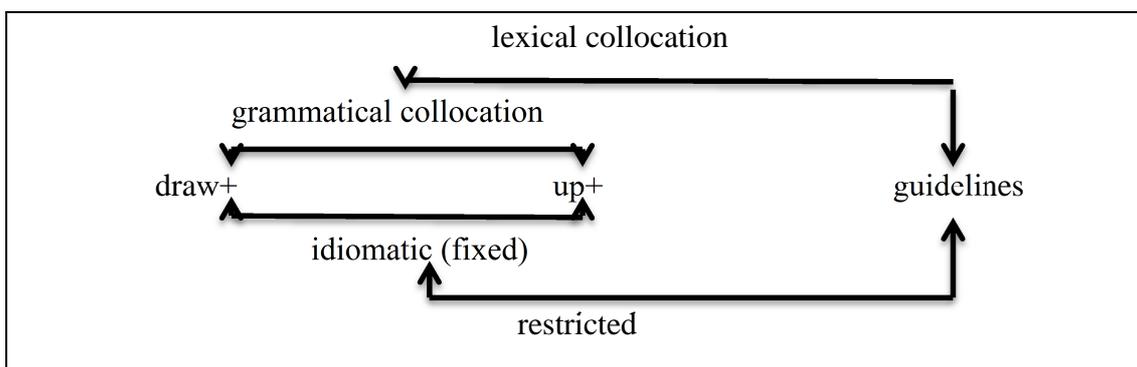
formulaic expressions and further to generate them, and to perceive the distribution of the salient words.

In the previous sections, I have explored the general characteristics of learner language, and teaching phraseologies; now the focus will be turned to PVs and learning.

### 3.4 Learning PVs

#### 3.4.1 Collocation and PVs

The linguistic environments of PVs are challenging to learners, for instance, collocations. A call for supplying information regarding the collocates of PVs has been made in studies such as Cowie (1993:41); therefore this section will relate the notion of collocation to phrasal verbs. The collocation scope of a PV can be illustrated in the case of a transitive PV *draw up* with *guidelines*:



**Figure 3.1: The example of “draw up guidelines”**

There are two levels of collocation: one is the internal structure of the PV and the other concerns the relation of the PV and its most relevant neighbour words (for example, in the VN collocation which is widely explored in the studies of phraseology, the noun is usually the direct object of the verb). The two combinational types, grammatical collocations and lexical collocations, are distinguished in Benson et al. (1986:191). The former refers to the combinations of noun/adjective + a closed class word (e.g. prepositions), and the latter consists of two open class words (e.g. V + N). Therefore, *draw up* can be seen as a grammatical collocation, while *draw up + guidelines* falls into the lexical type of collocation.

This view involves the consideration of word class, but the collocation relationship can also be classified by semantic transparency or the degrees of restriction (these two notions are discussed in depth in Section 6.4.1). The internal relationship within a PV is often accounted for by its semantic transparency (or opacity). Within the unit of a PV, the range goes from semantically transparent combinations such as *go out* to idiomatic PV as *turn up* (arrive at somewhere). Idiomatic PVs often share the properties of idioms, including ‘semantic opacity’ and ‘structural stability’. The meaning is not composed by the elements of the PV but conventionally assigned, and neither the verb nor the particle can be replaced by other words, otherwise the meaning changes radically. In this case,

*draw up*, the internal relationship is idiomatic or fixed. When it extends to the direct neighbouring word, many collocates are possible candidates (e.g. *guidelines/documents/outline*, etc.) although the number is limited. This is accounted for by degrees of restriction.

### **3.4.2 Problems of learning and teaching PVs**

PVs remain a major challenge for L2 learners because no really accurate description of them is available. They have been taken as a serious learning hurdle and many researchers have pointed out where the difficulties lie (McArthur, 1979, 1989). For example, De Cock (2005:16-18) summarises the common problems of PVs for learners: avoidance, style deficiency, semantic confusion, lack of collocational awareness, using idiosyncratic phrasal verbs and syntactic error. Avoidance is one of these problems that make PVs notorious for foreign language learners. PVs have been found to be 'avoided' by learners in many studies (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007). However, it is not possible to assign absence definitely to the strategy of avoidance in a corpus-based study, thus no further details will be discussed here.

Other studies conducted in light of learners' difficulty also reveal several general findings (Dagut & Laufer, 1985; Hägglund, 2001; Sjöhlom, 1995). Firstly, literal PVs

are widely preferred to idiomatic PVs across different L1 backgrounds. Secondly, learners tend to use PVs less frequently than do native speakers. Thirdly, the structural difference between L1-L2 (some languages do not have VPCs) will cause problems for learners. Moreover, difficulties may result from “polysemy, contextual and collocational restrictions, phrasal verb combinations, grammatical environment” (Lennon, 1996). The multiplicity of senses of PVs is also recognised as a hurdle for learners in Cornell (1985). Furthermore, from a didactic perspective, Side (1990:144-145) lists eight reasons for students’ resistance to learning PVs:

1. Confusion of combining the verb and the particle
2. Polysemy of PVs
3. Opacity of the meanings of idiomatic PVs
4. Preference of a synonymous latinized one-word verb to a two-word PV
5. The particle seems random
6. Confusion of transitivity and separability
7. Register/appropriacy
8. First language interference

Overall, it can be predicted that learners may have problems on two levels. The first is the collocation of the verb and the particle. They must know the correct combination

to express the right meaning and the idiomaticity (meaning opacity) of the PV. The second is the collocation of the PV and its direct collocates. The selection of appropriate collocates could become a difficulty for learners, whether they are semantically or arbitrarily determined. These two levels will be discussed in the following paragraphs.

For the first level, idiomaticity has been noticed as the marked semantic feature of phrasal verbs; however, this issue has not attracted attention in equal weight to its importance. Especially for learner language studies, idiomaticity is not a peripheral area in studies which focus on phrasal verbs, particularly when the problem of learnability is involved, as stated by Waibel (2007) :

*It is desirable to investigate this important aspect of phrasal verbs, especially in view of the fact that qualified statements about the learnability of phrasal verbs have to be based on a comparison of performance as regards transparent and idiomatic phrasal verbs. (Waibel, 2007:165)*

PVs have been found to be avoided or underused by learners of many first language backgrounds (Dagut & Laufer, 1985; Liao & Fukuya, 2004). In these studies, idiomaticity is usually taken as an important factor to explain why learners fail to be fully competent in using phrasal verbs. Most studies are based on the presumption that

more idiomatic/figurative PVs will result in more difficulty for learners. However, this may be true only when the learners are examined on their receptive knowledge. In other words, it makes sense to say that learners have no way to know an idiomatic PV which has never been encountered before, but once the sense of the PV is revealed to them, the meaning can be easily acquired. However, this is not enough; they also need to know the usages in different contexts.

The second level involves another problem of learning PVs: learners are not sensitive to the collocations, especially to those which are restricted to some degree. In other words, they have less difficulty using those which are extremely restricted or completely free (Howarth, 1998; Nesselhauf, 2005). An important finding of Nesselhauf's (2005) work is that it is not the most restricted collocations that are most difficult for learners but the combinations of less restriction, namely the less restricted collocations which the node word can take more collocates. The example of the more restricted combination given by Nesselhauf (2005) is *pay attention*, in contrast to the less restricted combination such as *perform*, which can collocate with *an experiment*, *a miracle*, *a ceremony*, etc. A semantic constraint specifies the conditions which its collocates have to satisfy. However, restricted collocations are problematic to L2

learners. We know that the collocates of the PV used in one sense are constrained in the same semantic field, but the semantic restrictions are hard to capture.

The discovery of these learning problems has led to a change in the teaching paradigm of PVs. As mentioned earlier, previous studies have mainly addressed the issue of the syntax problems, but only a little work has been undertaken empirically to explore the acquisition of syntactic rules (Sawyer, 1999). Regarding the teaching method, teaching grammatical rules and form-meaning mappings explicitly to learners has been shown to be effective (Gallagher, 2006; Thibeuau, 1999); however, the teaching of PVs has been criticised as involving an overemphasis on syntactic structures, so the inclusion of semantic features has been advocated instead (Dalle, 1983). Researchers have commonly proposed teaching PVs according to their regularities, fixedness or categorisation (Smidowicz, 1997). By doing so, their lexical nature is highlighted. As a result, some scholars have proposed teaching PVs using a lexical approach (see Lewis, 1993). Also, the importance of learning PVs by their contexts, semantic fields, etc. has been addressed by Klein (1995). The roles of particles and contexts have been noticed by Side (1990:151) as well, and he emphasises the importance of prioritising the particle of a PV and the need to put PVs in their

contextualisation. In addition, researchers often suggest learning PVs by using authentic texts (Wyss, 2003).

### **3.4.3 Other studies of PVs**

Phrasal verbs are interesting items which are considered a rich resource for shedding light on cognition and formulaic language. PVs are also an appealing area for cognitive linguists, who recommend that the meanings/functions of the particles be emphasised to facilitate learning PVs (De Rycker, 2005; Hannan, 1998; Kurtyka, 2001; Lindner, 1983; Neagu, 2007).

Other kind of phrasal verb studies concern formulaic language and are often conducted using a corpus approach. Those which investigate PVs can be divided into two purposes. The first is to probe PVs in the learner language and to identify the differences between groups of speakers (Hägglund, 2001; Waibel, 2007). This type of research is concerned with how PVs are used by the learners and the unnatural features that make them deviate from native norms. Among them, Waibel (2007:130) concludes three characteristics of the PV uses in the Italian and German learners' languages which lead to 'un-naturalness': (1) collocational deviations, (2) the inappropriate choice of a phrasal verb, and (3) the simplified use of phrasal verbs. A special type of research

studies issues related to PVs, such as Moon (2005), which deals with metaphors and PVs. Another common thread in PV research is to survey frequent PVs or a complete list of PVs in L1 English (Gardner & Davies, 2007; Kaalep & Muischnek, 2002). Several researchers have surveyed lexical verbs and particles which are used most frequently to form phrasal verbs. Their findings are listed in Table 3.1:

**Table 3.1: Most productive verbs and particles of phrasal verbs**

	Waibel (2007)	Biber et al. (1999)	Gardner et al. (2007)
Lexical verbs	bring	bring	go
	come	come	come
	find	get	take
	get	go	get
	give	put	set
	go	take	carry
	keep	set	turn
	make	turn	bring
	put		look
	take		put
Particles			.....
		down	out
		in	up
		off	down
		on	back
		out	off
		up	round
			.....

Note: The 2nd and 3rd columns are not ranked. The list in the 4th column is not complete, only some of the top words are listed.

It is noteworthy that these three studies have been conducted in order to cover the lexical verbs which are combined with each possible particle in the corpus. If we only look at verbs which are followed by one particle, say UP, the list of verbs will be somewhat different. For example, the verbs which Gardner and Davies (2007) find to combine with UP are: *set, pick, come, make, take, give, get, look, go, put* (in rank order).

### **3.5 Summary**

This chapter has reviewed the characteristics of learner language in the literature, indicating learners may tend to use a small range of general vocabulary and less formulaic units. It then can be conjectured, in terms of PV uses, that a similar situation will be found in this present study that the Chinese learners may rely on certain limited general or basic PVs, and they may produce less or untypical/atypical phraseologies. LL contains many non-native features which distinguish it from native language. The imperfection of LL is a consequence of learners' insufficient lexical knowledge. Among aspects of lexical knowledge, collocational knowledge is the top area of which learners are in need. Such a lack of knowledge can be improved by the assistance of introducing corpora applications to teaching.

Learner corpora have been demonstrated to be a good means to investigate

learner languages. Therefore a Chinese learner corpus (CLEC) will be probed in this study with comparison of a native corpus LOCNESS, both of which will be introduced in details in Chapter 5.

As the focus of this thesis is PVs, the relations of PVs and their collocations have been introduced, and the learning problems involved summarised. I have argued that the idiomaticity/opaqueness of a PV may not be an insurmountable obstacle for learners, but learners are faced with problems of selecting and combing relevant phraseologies appropriately. As a result, in the next chapter, issues of phraseology will be discussed in order to establish the theoretical grounds for this thesis.

# **Chapter4: PHRASEOLOGY**

## **4.1 Introduction**

This chapter reviews the concept of phraseology and the development of the relevant research. It also describes the relationship between meaning and phraseology, introducing ideas of major concern in this thesis, such as collocation, semantic preference, semantic sequence and the contextual approach, etc. This is followed by a section which presents the current state of affairs in the domain of phraseology in learner language studies. The need to explore learners' phraseological performance through a more flexible approach is brought to the fore throughout this chapter. This information is intended to contribute to the establishment of an adequate background for the remainder of the thesis.

The notion of phraseology is introduced in detail in Cowie (2005) from the Russian traditions to the present studies of collocations. The classic Russian school can be represented by the work of Vinogradov and Amosova (see Cowie, 2005:4ff), who focus on the description of phraseological classifications. Among the other strands of phraseological studies, the British neo-Firthian tradition is particularly relevant to this present work. The text analysis approach developed from Firth (1957a, 1957b) and

Halliday (1992) to Sinclair (1991, 1996, 2004b) has shed great light on a new perspective on meanings (see the summary in Stubbs, 1996:Ch2). Among the British neo-Firthian researchers, great impact on language learning comes from Sinclair's (1991) proposal of the principles of 'idiom' and 'open-choice', which control text organisation. In his view, language is made up of "a large number of semi-preconstructed phrases that constitute single choices" (ibid: 110). The 'open-choice principle' is that "words are treated as independent items of meaning", thus "each represents a separate choice" (Sinclair, 1991:175). The 'idiom principle' is that "the choice of one word affects the choice of others in its vicinity" (Sinclair, 1991:173). The latter is of particular importance to recent studies of phraseology. It is convincing that words are produced non-randomly in a great proportion of language, as semi-preconstructed phrases are observed to be pervasive. The idiom principle accounts for what the open-choice/grammar-based principle cannot explain.

The idea that language comprises a large amount of phraseological units has become widely recognised. The absence or under-representation of these phraseologies certainly makes learner language 'non-nativelike', but the presence of these restricted/semi-fixed constituents may also suggest the possible causes of the

‘un-naturalness’ in learner language if they are not like the native phraseologies. This will be revealed from studies of phraseologies in learner languages (see Section 4.3.5).

## **4.2 Phraseology and other related concepts**

### **4.2.1 Phraseology**

At the outset, before we can make use of the notion of phraseology, it is essential to clarify how this terminology is generally construed and the important elements that are involved. The concept ‘phraseology’ covers many different terms used by different researchers. Terms used for this concept include: formulas, ready-made language, extended units of meaning/lexical items, pattern grammar, lexical bundles, lexical phrases, clusters, n-grams, skipgrams, phrase-frames, phrasal constructions, phrasemes, prefabs, and recurrent word combinations, among (Weinert, 1995; Wray, 2002; Cheng, Greaves, Sinclair, & Warren, 2009).

These technical terms generally cover common ground but with some differences. For example, Biber’s (2009:282) lexical bundles emphasise that the recurrent sequences are in the scope of one register. N-grams often refer to contiguous word sequences automatically extracted by computer programs (Cheng, Greaves, & Warren, 2006; Granger & Paquot, 2008) while recurrent word combinations are adopted by

Altenberg (2005:101) to indicate a “continuous string of words occurring more than once in identical form”.

Besides these different technical terminologies, the term ‘phraseology’ is widely used in many studies. The meaning of phraseology, however, is not made clear in most literature whose target of discussion is phraseology itself. It is often taken for granted that everyone knows what phraseology means. However, in fact, different interpretations have been proposed by different scholars or in different fields. For instance, Teliya et al. (2005:55) sees phraseology as “a domain of linguistic study which to a high degree illustrates the correlation between language and culture”. Apart from such a particular viewpoint required in studies which have specific needs, the term phraseology is given a general and broad definition. For example, Cowie defines it as “the study of the structure, meaning and use of word combinations” (1994:3168). Gries (2008:4) also gives a broad definition of phraseology: “the co-occurrence of a form or a lemma of a lexical item and any other kind of linguistic elements (word/grammatical patterns)”. The tendency of future research seems to welcome a more general definition of phraseology, because it is taken as a superordinate term to encompass a number of phenomena.

Some researchers have attempted to set out the conditions that a phraseological unit has to meet. For example, Waibel (2007:5) summarises five such conditions, including their “multi-word character, lexicalisation, fixedness, institutionalisation and non-compositionality”. Amongst these, non-compositionality is not necessarily required, since it can be taken as a continuum with degrees of opaqueness.

### **4.2.2 Idiomaticity**

When discussing phraseology, one concept which has to be mentioned is ‘idiomaticity’. The term ‘idiomaticity’ is often used in two ways. The first use is similar to phraseology in language, and a short review will be presented shortly. The second is specifically used to refer to the opaqueness of phrasal verbs, therefore is considered in the discussion of phrasal verbs (see Chapter 2). Occasionally idiomaticity is adopted in particular to refer to “the psychological construct of a quality that speakers create on the basis of the different idiomatic variation parameters” (Wulff, 2008:4). The variation parameters are compositionality and syntactic flexibility (Wulff, 2008:4). An example is native speakers can distinguish the degrees of idiomaticity degrees between *take the plunge* and *write a letter* (Wulff, 2008:1). As this view of idiomaticity is not the main concern of this thesis, no further details will be mentioned. With a special concern for idiomaticity in language learning, Warren (2005:35) proposes two specific definitions

of idiomaticity: (1) “native selection of expressions (2) that which one has to know over and above rules and words”. Moreover, idiomaticity carries over from the phrase level to the clause level and the discourse level. At the phrase level, that is, when it comes to word combination, Warren points out that knowing a large number of idioms does not necessarily mean having great idiomaticity. Not only idioms but also other restricted combinations all contribute to feature idiomaticity. To achieve idiomaticity, a learner needs to know the collocational restrictions, which can be classified into requirements of certain ‘meaning’ or ‘item’. The former is exemplified by the combination *look forward to* with positive situations; the latter can be represented by instances such as *brush teeth* or *pull somebody’s leg* (Warren, 2005:42). From the model of idiomaticity, Warren establishes that performing idiomaticity involves knowing discourse structure, formal idioms and lexicalised sentence stems, expressions used in social interaction and the combinatory potentials of words.

### **4.2.3 Collocation**

Another term, ‘collocation’, is also of particular importance to phraseology because the two terms ‘collocation’ and ‘phraseology’ on some occasions overlap while on other occasions denote different notions. As has seen above (Cowie, 2005; Gries, 2008), phraseology usually refers to the discipline of research on ‘fixedness’ of language

systems (Howarth, 1996:6), but it can also describe the ‘co-occurrence’ of linguistic items. Similarly, collocation is often used as a technical term which covers the ‘co-occurrence’ of two or more lexical items (Poulsen, 2005:14), so it is often recognised as a synonym of phraseology or a subcategory of it. Although they are interchangeable to some extent, the term phraseology usually covers varieties of prefabricated units in a language, while the term collocation often represents the actual linguistic combinations and is mostly used in empirical studies.

There are two main different schools of thought behind the identification of phraseology/collocation: ‘the statistically orientated approach’ and ‘the significance orientated approach’, as identified by Herbst (1996:380). The former is also called ‘frequency-based approach’ (Nesselhauf, 2004), or ‘distributional approach’ (Granger & Paquot, 2008:27), and the latter is also termed as ‘phraseological approach’ (Nesselhauf, 2004). The reason of this division is that some researchers consider collocation as a pure statistical phenomenon or the co-occurrence of two or more items in a given span (e.g. Sinclair, 1991; Stubbs, 2002), while others take the view that collocation is phraseological (a type of word combination), thus minimising the dependence on frequency (Cowie, 1994; Nesselhauf, 2003:224). Granger and Paquot (2008) add that the difference is a result of some researchers focusing more on

‘fixedness’, while others focus on the less fixed ‘collocation’. The distributional approach includes n-grams and co-occurrence analysis, which separate continuous and discontinuous word combinations (Granger & Paquot, 2008:38-39). Biber (2009:276) also gives a summary that studies of phraseology can be distinguished either by the targets of research or the types of approaches. For research targets, attention could be drawn to idiomatic expressions (for example, *in a nutshell*) or salient multi-word sequences (for example, *you're never going to believe this*). For research approaches, a corpus-based study is conducted on the grounds of some preconceived linguistic theories. In contrast, a corpus-driven one assumes as little as possible such existing theories (for more discussion of corpus-based and corpus-driven approaches, (cf. Tognini-Bonelli, 2001: 65-100).

I will turn to focus on the issue of ‘collocation’, combination of words, which has not been a central issue of general linguistics until recent years. Interest in collocation derives from the studies of multi-word patterns (Cowie, 1998). Lately, attention has been drawn to collocation in learner languages (Howarth, 1996; Nesselhauf, 2005). Since collocation has been the core issue of phraseology studies, Hunston (2002:76-79) summarises the uses of collocations in the previous studies: (1) to highlight the different meanings of a word; (2) to obtain the dominant phraseology of a word; (3) to

obtain a profile of the semantic field of a word. This suggests that collocation plays a crucial role in studying the behaviours of a word, thus this will be one of the primary concerns in this thesis. In the following paragraphs, I will address the notion of collocation first and then narrow down to the analysis problems.

Collocation is introduced as an important issue by Firth (1957a, 1957b), who proposes that the meaning of a word is determined by its collocation. The idea of collocation can be fleshed out by its definitions proposed by a number of researchers. Palmer (1933:5) states that “a collocation is a succession of two or more words that must be learned as an integral whole, and not pieced together from its component parts”. This view suggests that collocation can be the combination of two lexical items or a longer word-sequence, such as a phrase. It lays emphasis on learning a collocation as an integral unit. In the same vein, Cowie states that collocations are “associations of two or more lexemes (or roots) recognized in and defined by their occurrence in a specific range of grammatical constructions” (Cowie, 1994:3169). Benson et al. (1986) give collocation a simpler explanation. They regard collocations as ‘recurrent combinations’, and group them into lexical and grammatical collocations. Stubbs (1995:23) uses collocation to mean “a relationship of habitual co-occurrence between words (lemmas or word-forms)”. Sometimes arbitrariness is emphasised, as in the definition by

Nesselfhauf (2005:1) that collocations are “arbitrarily restricted lexeme combinations”. Partington (1998:15-16) gives a clear summary of how the definitions of collocation develop. He starts with Firth’s famous statement, then adds Sinclair’s view that “collocation is the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991:170). Based on these definitions, we may distinguish groups of collocation: in one of the groups the constituents of a collocation of words have a semantic relationship and may form a lexical set, for instance *sheer volume/number/scale/size* (Partington, 1998:34-35). The connection of the words in a collocation may also be arbitrary, as demonstrated by the example *strong tea*. In another group, the collocation of words makes a formula or idiom such as *as well as* and *kick the bucket* (Moon, 1998:47).

For other researchers, further focus is placed on the relation between the elements. Nattinger and DeCarrico (1992:21) regard collocations as “strings of words that seem to have certain ‘mutual expectancy’, or a greater-than-chance likelihood that they will co-occur in any text”. This considers the collocational relation between the two related words and raises the question of direction (the differences from V to N and from N to V). For example, *commit* can be followed by some options such as *suicide/murder/crime*, but for the noun *suicide*, the verb expected is always *commit*.

This definition also stresses that collocation must exclude combinations which co-occur by chance. This problem can be easily fixed by using statistical tests (see Chapter 5). Besides these theoretical viewpoints, studies have taken a practical and pedagogical perspective, emphasising how a word is used with its accompanying words in real contexts. For example, Granger (2005a:146) considers collocation to be “the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its ‘synonyms’ because of constraints which are not on the level of syntax or conceptual meaning but on that of usage”. This viewpoint pays more attention to the collocations which are arbitrarily connected, thus are difficult for learners to predict.

Despite that definitions of collocation are provided by many researchers as seen above, some other researchers consider these definitions to be vague and hard to capture. For instance, Fontenelle points out that to describe collocations as “groups of words which frequently occur in combination with each other” is not enough. The definitions must include ‘number of elements’, ‘frequency of occurrence’ and ‘classes of words’ (Fontenelle, 2005:191). At the same time, Kennedy (1998:111) also warns that the criteria used to judge them are controversial, and he argues that questions such as those below must be solved before a genuine collocation can be located:

- Variations: *how many elements are fixed?*
- Number of occurrences: *how frequently does a combination have to occur?*
- Forms or meanings: *Is it determined by syntactic or semantic criteria?*
- *Does a combination have to be well-formed or canonical?*
- *What's the degrees of collocability?*
- *Can collocations be lemmatized?*

Similarly, Gries (2008:2) also lists six parameters to define co-occurrence or phraseology, including the nature, the number of elements, the number of occurrences, the allowed distance, the degree of flexibility, and the role of semantic unity and non-compositionality. Indeed, adding these three conditions to define collocation/phraseology may help to make the notion more concrete and clear.

However, at the same time, a call to relax the strict conditions which qualify collocation/phraseology has also emerged. Granger and Paquot (2008:33) comment that collocation is often referred to as “arbitrarily restricted combinations of lexical words”, and acknowledge that definitions of collocation vary in different studies (Granger & Paquot, 2008:35). In their conclusion, they state that the range of studying collocation should not be focused only on ‘fixedness’ and ‘semantic non-compositionality’ (Granger & Paquot, 2008:45). They point out that a number of scholars such as Siepmann (2006) have arrived at a conclusion to widen the range of collocation. From a lexicographer's view, Siepmann (2006:2) defines a collocation as “any holistic lexical, lexico-grammatical or semantic unit which exhibits minimal

recurrence within a particular discourse community”, entailing four subtypes: “colligation, collocation between lexemes or phrasemes, collocation between lexemes and contextual features, and collocation between contextual features”. These subtypes indicate the need to extend a wider scope of collocations. These researchers advocate allowing for more flexibility when studying collocation.

In my opinion, the two positions: describing collocation on well-defined conditions and allowing more possibilities to be included in the scope of collocation are of different usefulness. The former position aims to clarify the notion of collocation by listing the necessary information to provide the researchers with complete and unambiguous criteria. The latter position suggests that the researchers open their mind in order to avoid the loss of useful data. However, I agree more with the second position which allows broader range of possibilities, because there are various types of collocation/phraseology. If we are conservative in offering as many conditions as possible to define the targets, it will not benefit discovering more types of them. Therefore, in this present study, I will not probe collocation/phraseology with any *a priori* defining conditions as those mentioned above.

From a psychological perspective, Hoey (2005:5) proposes a new definition of collocation:

*A psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution.*

Collocation is regarded as the consequence of priming, a term which Hoey (2005) borrows from the field of Psychology. In Hoey's view, "every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word" (2005:13). Priming, in this sense, forms and accounts for the knowledge of a word and its relations to other words, meanings, pragmatic functions, grammatical functions and positions in sentences or texts, which an individual develops in his/her mind.

Hoey (2005:13) differentiates several types of priming. The major type deals with the co-selection of words, namely collocations, further classified into: lexical collocates, semantic association, pragmatic association, textual collocations and textual semantic collocations. In addition, the sequence of part of speech is also considered important in his theory. This type contains colligations and textual colligations. His theory will be reviewed in more detail in Section 4.3.2.

So far we have noticed a discrepancy in how the previous research defined and treated collocation. As well as the theoretical respect, many empirical problems have

also emerged from the process of analysing collocations and increased the effort involved in dealing with collocations. As warned by Kennedy (1998:114), many genuine collocations and MWUs are not contiguous and may spill outside the typical 4:4 window proposed by Sinclair (1991:117). Altenberg and Granger (2001) also point out that one of the deficits of an automatic analysis (e.g. WordSmith) of collocates is it may produce many results which are not true ‘constructional collocates’, i.e. the collocates do not have a direct relation to the node word, such as *make* and *argument* in “*The goal in this type of argument is to make the public aware of the truths...*” (Altenberg & Granger, 2001:187). The direct relation between the node word and its collocates needs to be confirmed. Bearing this in mind, I will carry out a careful manual analysis to locate the relevant collocates. The collocates of the PVs analysed in this thesis will be restricted to the nouns which have a direct relation to the PV, so the PV and the noun form a meaningful combination.

### **4.3 Phraseology and meaning**

As mentioned earlier, the exploration of phraseology began in the former Soviet Union and Eastern Europe from the 1940s (Cowie, 2005). In the early days, scholarly attention was paid to the typologies of phraseology. Later on, research interests were

directed to recurrent strings of words (from short word collocations to longer phrase-like or clause-like idioms, frames or formulae). These types of studies form the mainstream research of phraseology, which can be represented by the work of Granger (2005a), Altenberg (2005) and Howarth (2005), among others.

However, more recently, combinations which have slots that can be filled with not only words but also more abstract units have been acknowledged by certain researchers. These combinations are examined in notions such as ‘semantic preferences’, ‘semantic prosody’ (Sections 4.3.1.1 and 4.3.1.2), ‘lexical priming’ and ‘semantic sequences’ (Sections 4.3.2 and 4.3.3). Before these notions are introduced, it is important to consider how meanings are formed. Since our languages are mostly idiomatic/phraseological, words often come together and are not isolated. Therefore the view that meanings are produced by individual words is challenged. In next section, we will turn to the notion of units of meaning, proposed by Sinclair (2004b).

### **4.3.1 Units of meaning**

A breakthrough in the understanding of how meaning is constructed has assisted our understanding of the roles of words in a language. It is often thought that meaning is inherent in an individual word. This is true in most cases, but new meanings can be created by the combination of words on abstract levels. Sinclair (2004b:25) puts

forward an observation of language that “words enter into meaningful relations with other words around them”. In his research on 'units of meaning', Sinclair (2004b:24) describes these abstract levels as a unit termed ‘lexical item’/‘extended units of meaning’, with a structure of collocation, colligation, semantic preference and semantic prosody. He alters this term to ‘Meaning Shift Unit’ (MSU) three years later (Sinclair, 2007a, 2007b), and this is called a ‘model of phraseology’ by Stubbs (2009:131). Because the main concern of this thesis lies in the co-selections of words, meanings or concepts, some of these meaning levels are central to my thesis. In particular, the notions of ‘semantic preference’ and ‘semantic prosody’ thus will be reviewed individually in detail in the following sections.

#### ***4.3.1.1 Semantic preference***

One important subcategory of the ‘extended units of meaning’ is semantic preference. Partington (2004:150) comments that “semantic preference ... remains ... tied to ... collocation”, because semantic preference is determined by frequently collocated lexical items. It is defined by Sinclair (2004b:142) as “the restriction of regular co-occurrence to items which share a semantic feature”. Stubbs (2001:65) offers a more specific definition of semantic preference: “the relation, not between individual words, but between a lemma or word-form and a set of semantically related words”.

He notes that *large* often occurs with [quantities and sizes]. In another study, Partington (1998) also successfully identifies several semantic preferences of the word *sheer*, including [magnitude], [force], [persistence] and [strong emotion], etc. Research has been carried out into this issue in the fields such as Business English (Nelson, 2006) and other specialised discourses. However, although the importance of semantic preference cannot be underestimated, the investigation of this matter is still in its infancy, especially in the domain of learner language. The extent to which learners acquire knowledge of semantic preference remains unknown. It is unclear how learners select the semantically-preferred words in relation to the lexis in question.

#### ***4.3.1.2 Semantic prosody***

Another notion in the ‘units of meaning’ is semantic prosody, which has connections to semantic preference. This issue has been addressed in the work of Sinclair (1991, 1996, 2004b:173), Louw (1993) and Stubbs (1995), among others. Semantic prosody concerns how core words are interpreted functionally in relation to context (Sinclair, 2004b:34). In contrast to semantic preference, semantic prosody often refers to the pragmatic connotation which sometimes gives an evaluative meaning to the text. For example, Sinclair (1991) notes that *happen* and *set in* are associated with unpleasant

things; Stubbs (1995) probes the verb *cause*, and discovers a similar result that it habitually comes with an unfavourable prosody. Partington (2004) takes a step further to study the synonyms of *happen* and *set in*, such as *occur*, *come about* and *take place*. Although it is partially true that words of the same semantic group share the same semantic prosody, these verbs are also found to have their own specific usages which cannot be accounted for by simply judging them as ‘negative’. Another study conducted by Kennedy (2008) bears witness to Partington’s conclusion. Kennedy investigates eight frequent verbs and found that even semantically close verbs may be profiled by different semantic prosody.

However, the notion of semantic prosody may not be as clear as it has been defined, because different researchers hold different views. Hunston (2007:250-258) pins down the differences between Sinclair and Partington’s perspectives to the extent of the elements which contribute to one semantic prosody. Sinclair (2004b) considers the whole ‘units of meaning’ as the constituents of semantic prosody, whereas Partington (2004) ascribes the prosody to the node word alone. It seems to be more justified to deduce a semantic prosody from the phraseology but not the word only, as evidenced by Hunston's (2007:254) examples of *persistent* and *persistence*. The idea that the phraseological units of a word come to form a ‘discourse function’ (i.e.

semantic prosody) sheds great light on the idea that phraseologies rather than word co-occurrence may be more accurate.

The difference between semantic preference and semantic prosody is also significant. Partington (2004:149-151) expounds this by citing a number of examples, coming to the conclusion that “semantic preferences combine to form an overriding prosody”, while semantic prosody “dictates the general environment which constrains the preferential choices”. This distinction is aimed to clarify the roles of these two notions.

### **4.3.2 Semantic/pragmatic associations and lexical priming**

A notion in strong relation to semantic preference and prosody is semantic association, which is termed by Hoey (2005). In the same book, he proposes the theory of ‘lexical priming’ to describe the tendency of words, meanings, grammatical configurations and textual positions, etc. to co-occur. Each time an individual encounters the co-occurrences, the strength of the association becomes stronger. Hoey’s work successfully substantiates the existence of various patterns/collocations at the lexical, syntactic and textual levels. The phenomena Hoey lays out in this book include: collocates, semantic associations, pragmatic associations, colligations, grammatical

categories, textual collocations, textual semantic associations and textual colligations.

Other than textual collocations, two types of association identified by Hoey (2005) are of particular relevance to my study: semantic and pragmatic associations. The former is pertinent to the general attributes shown by a group of words (usually of the same grammatical category) which co-occur with a node word. For example, Hoey (2005:25-26) studies the adjectives which precede the word *consequence*, and classifies them into four groups of semantic associations: [logic] (59%), [negative evaluation] (15%), [seriousness] (11%) and [unexpectedness] (6%). The latter accounts for the phenomenon where “a word or word sequence associated with a set of features all serve the same or similar pragmatic functions” (Hoey, 2005:26). The illustrations are the co-text in respect to the word *sixty*, which indicates [vagueness]; and *reason*, which is found to associate with [acts of denial].

Hoey explains that the term ‘semantic association’ is adopted in his book instead of ‘semantic prosody’ because of the controversy it raises (see above). But Hoey regards the term ‘semantic association’ to be similar to Sinclair’s (1999) ‘semantic preference’. Semantic association is a set of ideas at a higher level than lexical collocations. It is the assemblage of shared features extracted from a group of collocates. The employment of semantic association helps to display the regularity of

language.

One of the insightful ideas in Hoey's (2005:16) book is that he points out that semantic association cannot be discerned by probing collocations alone, especially if only specific collocates are examined. He states: "the priming is operating at a more abstract level". The typical example given by Hoey is *thirty hour ride*, which can be represented as [number]-[time]-[journey]. He argues it is not the word *thirty* which consistently co-occurs with *hour* that matters. What is significant is the fact that the abstract semantic units [number] and [time] are primed together. This combination shows clearly how the lexical collocation can be elevated to a higher level where each semantic association label reflects a semantic domain. Hoey's proposal of semantic association chimes with Sinclair's view of semantic preference, that meaning patterns can be corroborated at a more abstract/general level.

Now let us consider the other association type: pragmatic association, which Hoey (2005:27) illustrates using the example of the word *sixty*. He finds many instances from his corpus, such as *about sixty*, *over sixty*, *more than sixty*, *an average of sixty*, *fifty to sixty*, *sixty or more*, *sixty-some*. As seen, no recurrent or repetitive word can be discerned in the context of *sixty*, but in all of these cases, they create a pragmatic function of [vagueness]. An important implication of this example is that an

abstract idea such as [vagueness] can be expressed by diverse forms. The collocations are not necessarily limited to combinations of words; word sequences/phrases or other structures are also possible.

Hoey (2005:31) also points out that semantic association has a “tight syntactic relationship” while pragmatic association does not. Semantic association is found to be restricted grammatically in Hoey's data. For example, the collocates of *consequence* are consistently found in the position immediately prior to it and they are all adjectives. This restriction does not apply to pragmatic association, as illustrated by the instances of *sixty* above. So, these two notions seem to have distinctive properties. Semantic association explains the relationship of lexical words, which is a generalisation of the possible collocates in an assigned position. Pragmatic association describes the pragmatic function which can be performed by words or word sequences. However, although Hoey makes a differentiation between these two types of association, he admits that their distinction is blurred because the effect of priming is the same.

### **4.3.3 Semantic sequence**

As we have seen, the majority of studies concentrate on one subcategory of the phraseology. It is good to focus on one specific phenomenon, but this also has the

downside of neglecting the fact that many subcategories are mixed, inseparable and co-existent. From the perspective of usage, these subcategories seldom occur alone. Most of the time, language learners acquire them in a package.

As a result, an innovative term, semantic sequence, is proposed by Hunston (2008) to refer to hybrid phraseological units which may encompass the combination of all possible subcategories. She states they are “series of meaning elements”, in other words, ‘lexic(o)(al)-grammatical patterns’ (Altenberg & Granger, 2001; Nesselhauf & Römer, 2007), composed of fixed or restricted words or phrases, accompanied by variable slots. Table 4.1 presents the semantic sequence examples found by Charles (2004), Groom (2007) and Hunston (2008). In Hunston’s study (2008:277), a semantic sequence like ‘[possibility] + *to make sure* + *that*-clause’ has a semantic field, a core phrase and a grammatical pattern. As seen, a semantic sequence would mainly be constituted by elements of a few specific words/phrases and a number of meanings/semantic fields, and could also be mixed with others such as clause types. It should be noted that the semantic or meaning element is not a well-delimited concept: sometimes it may be represented by a set of ‘semantic labels’ (e.g. volition, obligation, possibility) or ‘discourse functions’ (e.g. the idea, suggestion exists/is evaluated/causes something), as identified by Hunston (2008). The idea of

semantic sequences is close to Schmitt and Carter's (2004:7) 'flexible formulaic sequence' which consists of some fixed elements and some flexible slots. These slots, like the labels of semantic description in semantic sequences, are subject to semantic constraints. Meanwhile Schmitt and Carter warn that the amount of these sequences may be much larger than is estimated by the present software. Their significance is widespread in any language, thus they will be the key to foreign language learning. Given that semantic sequences are hard to be detected directly by computer programs from corpora data because they can be realised by a diversity of forms. The only way to discern them so far is by human analysis, where the researcher digests the texts and transforms them into a series of meaning elements.

**Table 4.1: Examples of semantic sequences**

Author	Semantic sequence
Charles (2004)  (2011)	[logical basis] + <i>it is clear that</i> + [claim] <i>although/though</i> + [positive evaluation] + [negative evaluation] + [reason] (optional)
Groom (2007)	[entity] + [existence etc.] + <i>beyond</i> + [conventional] + <i>to</i> + [new domain] [statistical indicators] + <i>among</i> + [social group or institution] [relationship] + <i>among</i> + [conceptual phenomena]
Hunston (2008)	[possibility] + <i>to make sure</i> + <i>that</i> -clause [theory/argument] + [arise from] + <i>the observation</i> + <i>that</i> -clause <i>the observation</i> + <i>that</i> -clause + [consistency] + [theory/argument]

This concept seems to overlap with others such as ‘units of meaning’, ‘lexical bundles’, ‘lexical priming’ and ‘construction grammar’, etc. Hunston (2008:298) discusses the distinctions between semantic sequence and these other terms. Despite some small differences, they actually share considerable commonalities. The diversity of terms is the result of tackling a problem from different angles. Semantic sequences are less theoretically robust because they are “the product of observation”, as pointed out by Hunston (2008:298). Thus in this study, I will take ‘semantic sequence’ as a concept which allows for more flexibility and variation.

As Hunston (2008:272,284) notes, semantic sequences are more easily observed in specialised corpora. If learner language is taken as a special type of language, the texts produced by authors of the same first language background should exhibit consistent and systematic features which make up a specific corpus. That is to say, it may be fruitful to discover learner-specific semantic sequences in a learner corpus.

A search of the phraseology studies of learner language reveals that the phenomenon of ‘semantic sequence’ has been under-researched, demonstrating that the importance of semantic sequence has not been widely acknowledged. Therefore, this present work intends to contribute to exploring this gap in the literature.

#### 4.3.4 Contextual approach

As seen from the above reviewed research, the corpus approach adopted by the neo-Firthian researchers/Birmingham School, which takes into account all of the elements involved with a relationship is an appropriate one to obtain fruitful results. It is also regarded as benefiting L2 learners, as shown in Wible (2008:172), who advocates taking a contextual/Firthian perspective to introduce learners to an adequate lexical knowledge of the words they wish to learn.

Such an approach to take into account all the relevant lexical-grammatical information is called a 'contextual approach', and has been adopted in a number of research projects. For example, Sinclair (1991) investigates the phrasal verb *set in* and looked at several features relevant to this phrasal verb, including sentence length, level of the clause, position, word forms and subjects (see also Section 5.1). Following Mindt's (2000) idea of didactic grammar, Römer (2005) examines progressives by their function and context features. In her context feature analysis, she studies several 'context features': tense form, TO BE contraction, subject, preposition, object, time and place adverbial, negation, etc. A similar approach can be found in Moon (1998), who analyses the lexical and grammatical forms of formulaic expressions. Besides frequencies, she also examines grammatical types, inflectability,

and regular slots. In the category of grammatical types, she lists the syntactic roles which formulaic expressions and idioms (FEI) may have: predicate, nominal groups, predicative adjectival groups, modifiers, adjuncts, sentence adverbials, subordinate clauses and other classes. For inflectability, she tackles the problem of tense and mood. Furthermore, she probes the regular slots, which include: subject slots, non-subject slots, possessives, open slots. The grammatical types deal with the clausal positions of those FEIs, while the regular slots indicate co-selections between the FEI and its context neighbours. These studies above adopt a contextual approach which involves more grammatical concerns, but this present thesis will examine the relevant contextual features (i.e. the phraseological levels introduced above) where possible. This present study will adopt a contextual approach which focuses on the phraseological units but excludes many of the factors used by the work above, such as sentence length, positions, etc.

#### **4.3.5 Phraseology/collocation in learner language**

In Section 3.3.2, we have already seen the challenges phraseology poses in language teaching. This section will describe the roles of phraseology in learner language, including how phraseologies are acquired and performed. In addition, I will also discuss the findings from some significant studies which particularly concern

collocations.

The role of phraseology has been agreed to lie at the core of successful language learning. For instance, Wong-Fillmore (1976) notices the importance of acquiring formulaic speech in language learning. The reasons why phraseology plays a crucial role in learner language include: (1) phraseological units are a very common phenomenon; (2) phraseology has special functions. The first reason is evidenced by Howarth's (1996) estimation that in his corpus at least one third of the V+N combinations are collocations. The second reason is shown in Nesselhauf (2005:2). She summarises the functions of phraseological units from many researchers' work: They contribute to the production of creative language and fluency, facilitating comprehension and improving the users' similarity/likeness as a linguistic group.

With respect to the learning of phraseology, foreign or second language learners have a learning process which is very different from that which native speakers experience. It has been observed that the foreign language learners acquire English mainly from written input, in contrast to native speakers, who receive spoken input in most occasions (Wible, 2008). Learners of English have to deal with several problems such as unawareness of native-like selections and the opaqueness/restrictions of phraseologies (Waibel, 2007:7).

There is psychological and neurological evidence supporting the existence of formulaic language (see the summary by Weinert, 1995:185). The use of phraseologies, formulaic languages or prefabricated routines has already been observed widely in many first language studies, but they are also found present in L2. Foreign language learners have been observed to utilise formulaic units as strategies to achieve their communicative goals (Wray, 2002:178-183). They make use of formulaic wholes in the earlier stages and progress into analysed elements later (Nattinger & DeCarrico, 1992; Weinert, 1995; Schmitt & Carter, 2004).

Although L2 learners advance in proficiency, they fail to master the formulaicity of the target language. As noted by Yorio (1989), they make errors by inappropriately clustering words together; this was ascribed to their lack of competence in recognising the restrictions imposed on strings of words. Other research reports that L2 learners combine collocations, which are not natural co-occurrences or under-use formulaic units (Dechert & Lennon, 1989; Granger, 2005a). Therefore, L2 learners may achieve 'native-like fluency', but they still lack the ability of 'native-like selection' (Pawley & Syder, 1983). To conclude, unlike L1 learners who acquire languages with large units in the initial stages, L2 learners start with individual words, leading to difficulties in combining 'idiomatic' strings. They may produce grammatically possible but

non-idiomatic sequences because they have ‘too much choice’ (Wray, 2002:206).

Interest in phraseologies in learner languages has led to a surge of work on collocations or the grammatical-lexical patterns of one specific verb, such as that conducted by Altenberg and Granger (2001). Below I will focus on studies of collocations in learner language which are corpus-based in principle. Collocations, especially the restricted collocations in learner language, have been scrutinized by researchers such as Howarth (1996, 1998, 2005), Nesselhauf (2003, 2005), and Cross and Papp (2008), among many others. These studies have mainly investigated the collocations of the V + N constructions, with a focus on anomalous or erroneous combinations. General findings on collocation, with particular relevance to restrictions, are summarised in Nesselhauf (2005:8), where she concludes that learners tend to fall prey to restrictions which control what can and cannot be combined, at the same time. Other individual results often display specific learner performance, for example, Cross and Papp (2008:26) examine a Chinese English corpus and find that Chinese learners show a larger error rate in the use of verbs with prepositions in comparison with Greek and German learners.

The studies of collocations (two-word combination) undertaken by Howarth (1996, 2005) and Nesselhauf (2003, 2005) are most notable. Howarth (1996)

investigates the verb+noun collocations in one register (academic writing) produced by native and non-native writers of various L1s. His results show that learners make less use of restricted collocations than do native writers. He also claims that learners' general proficiency levels are not in correlation with their collocation performance. In another study, Howarth (2005:177) presents a similar report, where he states that "native speakers employ about 50 per cent more restricted collocations and idioms... than learners do, on average". Learners are found to be able to manage the idioms and free combinations (both ends of the idiomaticity spectrum) at an advanced level, but they are less aware of the mechanisms of the restricted collocations which lie in between. To reach native standard, Howarth concludes, the learners thus must be able to choose appropriate grammatical and lexical items, and select conventional collocations.

Nesselhauf (2003) examines the verb-noun collocations in the German subcorpus of ICLE produced by advanced English learners. The non-standard collocations are singled out by native speakers, and judged as to their acceptability. By doing so, the mistake types can be identified and the impact of the restriction of a collocation can be detected. The most common mistake types found are the wrong choice of verb and noun, as expected. The most interesting finding is that learners make most mistakes

with those collocations which have a medium degree of restriction, that is, “the verb takes a wider range of nouns” (Nesselhauf, 2003:233). In addition, the role of their first language is demonstrated to have great impact on their use of collocations. To make a full presentation of her study, Nesselhauf (2005) publishes a book which discusses the findings at length. Overall, the learners are found to be in line with Kaszubski’s (2000) finding that they tend to use fewer restricted collocations in comparison with native users. However, they also overuse some types of collocation that are deemed ‘safe’.

With an aim to explore the proportion of phraseological units in the native and non-native writings, Kaszubski’s (2000) thesis reports on Polish learners’ idiomatic performance based on six core verbs (*be, do, have, make, take, give*). The collocations are divided into groups of frozen, restricted and free combinations. According to her analysis, learners’ language is characterised by less use of idiomatic expressions than that of native users. The more advanced the learners are, the more idiomatic collocations can be found (Kaszubski, 2000:243). Another study undertaken by Wiktorsson (2002) on Swedish learners arrives at the same conclusion: the more proficient learners can display a higher quantity of multi-word combinations.

Besides the popular combinations of the VN structure, other linguists have

chosen to study collocations such as adverbs. For instance, Granger (2005a) studies amplifiers which end with *-ly* and function as modifiers, such as *perfectly natural*. In general, fewer amplifiers are used by French-speaking learners, and some of them, such as *completely, totally* are overused by the learners as ‘safe bets’, when compared with the natives (Granger, 2005a:148). Among the learners’ collocations, the majority are not native-like; in other words, they are never used by native speakers. She further categorises them into two sub-groups: maximisers (e.g. *totally, entirely*) and boosters (e.g. *highly, strongly*). There is not much difference in quantity found in the group of maximisers, whereas the boosters are used far more by the natives. Above all, a strong influence of the learners’ L1 is revealed in her data: the congruent collocations (the English combination has a direct translation equivalent in French) are found to be the learners' favourite choices. The equivalent in their L1 naturally enhances the learners’ confidence in making use of its English counterpart.

Because these above-mentioned studies all focus on the product of learner language, none of them concerns the phraseology in the input for the learners to learn. Durrant (2008) conducts lab-based and corpus-based research in his thesis to tackle this problem. In the lab-based study, the subjects (learners) are trained by reading texts which contain the target collocations in various times of recurrences. Then they

do a recall test to examine their memory retention. It is concluded that the learners are able to pick up and ‘learn’ these collocations from the input they were exposed to. The second, corpus-based, approach yields evidence that learners are more conservative in using new, coined collocations (Durrant, 2008:174). In accordance with the previous studies, the results also demonstrate that learners can use as many collocations as natives, but they tend to rely on some favoured types of collocation. Another crucial finding is that the non-idiomaticity of the non-native language may be a consequence of the learners’ lack of the “lower-frequency but strongly-associated” collocations (Durrant, 2008:183), but not the high-frequency collocations, as generally assumed.

Guo’s (2006:196-220) research also provides implications for this present study. His research is based on COLEC, a subcorpus of CLEC. COLEC consists of only examination essays while CLEC comprises more genres. Both corpora collect writings by Chinese students at similar level (from middle school to university), but differ slightly in that COLEC does not include writings from English-major students (Guo, 2006). He looks into the context of the combinations such as *take action*, *take place*, and *take on*. Since my thesis aims to probe phrasal verbs, his analysis of *take on* is of particular relevance here and worth some discussion. It is found that only some of the possible senses are employed by learners, thus revealing that the Chinese

learners do not have the full range of the usage of this PV at their disposal. He concludes that, as has been suggested by many previous studies, natives have the advantage of using more varied types of collocation, but learners are characterised by their limited and repeated uses of a few collocations (Guo, 2006:217).

#### **4.4 Summary**

As the previous sections illustrate, the terminologies that pertain to phraseology have been very complex and somewhat hinder the reconciliation of different theories. Although I agree that providing the basic information of the MWU in question is necessary, I feel that giving each term a rigorous definition does not help to clarify the fuzziness; on the contrary, it sometimes adds more opaqueness. Moreover, more recent research has suggested that many newly discovered phraseological units should be subsumed in the range of phraseology to increase its breadth of coverage, against the traditional view which is conservative in delimiting phraseological units. Therefore, in this thesis, many sub-terms are used interchangeably to avoid confusion, and at the same time allow more possibilities to be uncovered.

The notion of phraseology can be summarised by two important synopses given by Cowie (2005:12). He concludes that:

*Prefabricated expressions pervade all levels of linguistic organization--lexical, grammatical, pragmatic—and affect all kinds of structures, from entire utterances to simple phrases ... there are relatively few examples that are completely invariable or opaque.*

His observation provides intriguing implications for the study of phraseology.

First, since the phraseological elements are not limited in their function and size, the borderline of investigation can be broken, and the main concern of phraseology in language studies can extend across levels. Secondly, the flexibility/variability of certain slots in a phraseological unit should require more consideration than previously thought.

These implications are reflected in the work led by Sinclair, and followed by Hunston and Francis and others. Studies of phraseology such as Sinclair's (1991, 1996, 1999) and 'pattern grammar' (Hunston & Francis, 1999) have shown that there are systematic regularities of words, grammar and even meanings which display predictable patterns in a language. Their view of phraseology is not narrow: non-compositionality is not an essential criterion to define phraseology. Moreover, the parts that contribute to forming one meaning are taken as a group which can comprise linguistic items above the word level. Based on the results of these previous studies,

this present thesis makes an attempt to explore whether the regularities of MWUs can be found in learner language and to investigate whether learner language has the same patterns in regard to one MWU in the native English.

In the field of learner language, the role of phraseology has been centralised with the development of corpus studies. Nonetheless, scant attention has been paid to the issues of associations/co-selections. As a result, this thesis is situated within a broader scope of phraseology. It considers the commonly studied collocations, and also extends to the more abstract notions such as multiple-word combinations, semantic sequences (Hunston, 2008) and semantic associations (Hoey, 2005), etc. The examination of these phenomena is believed to manifest the discrepancy of idiomaticity in the learner language, and will help the researchers to capture more learner-specific characteristics.

# **Chapter5: METHODOLOGY**

## **5.1 Preliminaries**

This chapter introduces the methods of analysis used to examine the corpus data. Because different methods were employed in the analysis process, this present chapter will only report the general methodology taken throughout the research, that is, the extraction of phrasal verbs and their phraseologies. The different approaches to analysis which were adapted to meet the needs of the research purposes will be discussed at length later in the chapters on different particles where appropriate (see Chapters 6-8).

The methodology takes the ‘hypothesis-finding’ approach (Granger, 1998a) to probe PVs using a combination of quantitative and qualitative analysis. Accordingly, each chapter of the results will begin with the frequencies and distribution of the verbs which co-occur with one particle. Another key role in the methodology is the ‘contextual approach’ (Sinclair, 1991), which examines a lexical item in its related linguistic environment. Also based on the Contrastive Interlanguage Analysis (CIA) framework, the data analysis method is mainly a comparison of the native-speaker corpus and non-native-speaker corpus.

Not all of the aspects of usage knowledge pertaining to PVs can be fully revealed by a conventional error-analysis, but more information can be provided by a contextual approach. The term ‘contextual approach’ (cf. Chapter 4) is adopted by Römer (2005:5) to refer to a research approach which follows Firth and Sinclair’s ideas and techniques: it looks at the environmental elements that surround a node-word, including all of the phraseological units mentioned before. Such a holistic approach will start from probing a core word (which is a phrasal verb in this thesis), and then identify the words that habitually surround it. Further analysis can be undertaken by observing the semantic or syntactic patterns and extrapolating the more abstract notions such as semantic sequence and semantic prosody from the patterns.

Several researchers have carried out studies to ascertain the ‘lexical-grammatical’ patterns of a particular word or structure. For instance, the verb *make* was explored through its overuses and underuses, the categories of use (e.g. delexical or causative), and collocation (Altenberg & Granger, 2001). With the aim of exploiting learners’ lexical knowledge, the same verb *make* is also researched by Liu and Shaw (2001). They look at the similar lexical or grammatical factors and also other factors such as type-token ratios and prefabricated combinations. Besides single words, structures such as ‘future time progressives’ have also been studied using such a contextual approach,

including verb preferences, adverbial co-selections, subject types and negation (Nesselhauf & Römer, 2007:297). Regarding studying PVs, a similar analysis can also be found in Sinclair (1991:71), where he demonstrates an investigation into the patterns of phrasal verbs which are led by SET, such as *set about*, *set in*, *set off*. He observes phrasal verbs from several angles: sentence length, level of the clause (e.g. main or subordinate), and position of the PV in a sentence. Word-forms, or their tenses shown by the inflection of the verb, are also examined. Besides this, the subject of the PV in the sentence is also found to make a huge contribution to the meaning of the PV. Furthermore, his analysis also covers the notion of semantic prosody, which implies that PVs might be positive or negative, desirable or unpleasant (Sinclair, 1991:73-75).

In general there is a lack of studies which take on contextual associations and pose questions about learners' actual knowledge of one PV. We are unable to know the extent to which learners possess the same lexical knowledge as the native speakers. To close the gap, this study is conducted using analysis of actual uses of PVs. It is reasonable to contend that the language competence of learners should be explained at as many levels of phraseological units as possible. Therefore, the contextual approach I used to analyse the corpus data will start with the lexical and grammatical patterns of the PVs, and then extend to other phraseologies. The issues pertinent to PVs will be

approached in this thesis from two main dimensions: the properties of PVs and their relations to neighbour words.

The first research concern is the distribution of the PVs, types and tokens, and their overuses and underuses. By distribution I refer to the frequencies of each PV. They will be compared across corpora in order to obtain an overview of the quantitatively different uses by NSs and NNSs. For further analysis, overuses and underuses of PVs will be reported for some PVs, with an attempt to discover the disparities. Additionally, the productivity of the verb types and particle types will be compared in order to gain a clearer picture of the overview of PV uses. Lists of all the verbs and particles will also be made (see Appendices). The distribution section examines the frequencies of each PV, and this productivity section concentrates on the verbs and particles which are preferably used to construct PVs by NSs and NNSs. The type-token ratios will also be examined to uncover the disparity between NS and NNS corpora in conjunction with singling out the most frequent PVs. Meanwhile, preferences for word-forms and structures will also be probed for some PVs.

The second concern of this study is the primary goal of this thesis: to view PVs in a larger context. The patterns of PVs will be shown in concordance lines, and by sorting the concordance lines in different ways, hidden patterns can be highlighted and

compared. The patterns of the phraseological units will be carefully studied. These two dimensions are believed to manifest the differentiation of the subtle usages displayed by the NSs and NNSs.

In the next sections, the corpora and tools used are introduced first, followed by the data processing procedure. The data collected from the corpora are processed through two stages: extraction of PVs and extraction of phraseological units, both of which will be explained step by step below.

## **5.2 Corpora and tools**

### **5.2.1 Corpora selection and structure**

Based on the CIA (Contrastive Interlanguage Analysis) framework established by one-on-one or multiple NS versus NNS comparisons, this thesis compared a Chinese learner corpus with a native speaker one. The corpus searches were carried out on two corpora: a Chinese learner corpus, CLEC (Chinese Learner English Corpus), and a reference corpus, LOCNESS (Louvain Corpus of Native English Essays). CLEC is a Chinese learner corpus composed of five levels of learner attainment. The size of CLEC is about three times larger than LOCNESS, but the effect of this difference can

be minimised by normalisation. The CLEC is error-tagged by English teachers, and the texts are collected from free compositions and examination essays written by Chinese students (Gui & Yang, 2002). The topics of these articles in each subcorpus of CLEC are listed in Table 5.1 below:

**Table 5.1: Topics in CLEC**

<b>Level</b>	<b>Topics</b>
ST2	A Healthy Diet, The Most Impressive Thing in My Life, A party, A Day in My Weekend, An Interesting Story, A Very Nice Country, A Shop, My Hometown, A Letter, My Diary, Mid-Autumn Day
ST3	Getting to Know the World Outside the Campus, Practice Makes Perfect, Health Gains in Developing Countries, Global Shortage of Fresh Water, How to Make Good Use of College Life, My Bedroom, Social Activities and Our Study, Failure
ST4	Haste Makes Waste, My View on Job-Hopping, My View on Fake Commodities, Type of Speech, Friendship, Movies, Charm, A Chinese Holiday, Trust, Health Gains in Developing Countries, How to Make Dumplings
ST5	My Education, Chinese Festivals, My Grandmother, Galileo, The Use of Computer in the Modern World, A letter, A Diary
ST6	Euthanasia Should be Legalized in China, Prison System, Financial Reward, Television, The Military Service System, Peace, Equality, Crime

LOCNESS consists of argumentative and literary essays written by British and American native speakers. The British essays were written by A-level and university students; the American texts were all produced by university students, but are divided

into two types of essay: argumentative and literary-mixed. The topics of the A-level essays include: National Lottery, Computers and the Human Brain, BSE and Eating Beef, etc. The British university students write about topics such as French Intellectual Tradition, or A Single Europe: A Loss of Sovereignty for Britain. The American argumentative essays have topics such as: Euthanasia, Sex Equality, Ethics, Suicide, Crime Does Not Pay, The Welfare System, Divorce, Death Penalty, Rules and Regulations, Drinking Age, Sink or Swim, etc. The American literary-mixed essays were produced on topics such as: Who is Hamlet? and Aspects of Social Psychology, etc. The full details of the components can be found online (see the website of LOCNESS).

For the native corpus, LOCNESS is used. Although LOCNESS is not strictly comparable with CLEC (see discussion below), many researchers use it in comparison with learner corpora, because it is publicly available and its writers are all college students who share a similar age and background with those recorded in CLEC. The native students' writing is considered as the appropriate target language which the student learners are applying themselves to achieve. The large-scale corpus, BoE, is used for consultation when the comparison cannot be made due to absence of data and the needs to set up standards emerge.

The structures of the two corpora, CLEC and LOCNESS, are described below to investigate their comparability. The word counts, writer backgrounds and essay types are shown in Table 5.2 and Table 5.3.

**Table 5.2: Structure and size of CLEC**

CLEC	N of words
<b>Five subcorpora</b>	
ST2 (senior high school students)	208,088
ST3(first/second year college students—non English major)	209,043
ST4(third/fourth year college students—non English major)	212,855
ST5(first/second year college students—English major)	214,510
ST6(third/fourth year college students—English major)	226,106
TOTAL	1,070,602

**Table 5.3: Structure and size of LOCNESS**

LOCNESS	N of words
argumentative essays(American university students)	149,574
literary-mixed essays(American university students)	18,826
argumentative and literary essays (British university students)	95,695
British A-level argumentative essays	60,209
TOTAL	324,304

As CLEC is composed of five subcorpora which belong to different proficiency levels, the reason to mix them together as a representation of the Chinese learner language needs to be explained. Guo, in his study (2006) examines a Chinese learner corpus COLEC (the former version of CLEC), and he combines two levels of the

Chinese learners due to their large extent of similarities such as ‘culture, learning purposes, consistency of errors’, etc., recognized by several researchers (see *ibid*: 51).

For the same reason, I will follow Guo’s method to take these subcorpora as a whole, and also for the purpose to keep as much data as possible since PVs are not extremely frequent targets.

The comparability of these two corpora requires some discussion here concerning the sizes and genres of the collections of texts. The first problem is the simpler one and can be solved easily: given that these two corpora are of different sizes, the frequencies of the target PVs should be subjected to normalisation. By doing so, they can be said to be comparable from the statistical point of view.

The second problem is trickier, because although the texts of LOCNESS are mainly argumentative essays, the Chinese learner corpus is made up of a lower proportion of such a type of genre. On top of that, the topics assigned to the Chinese students and the native students are different. There may be a ‘topic effect’ which influences the phrasal verb selection for both the NS and NNS writers. It can be argued that the writers will stick to one or more specific PV types when they discuss one special topic. Guo (2006) finds that some vocabulary differences are due to the small number of topics in his Chinese learner corpus COLEC. Besides, the exact numbers of texts with particular

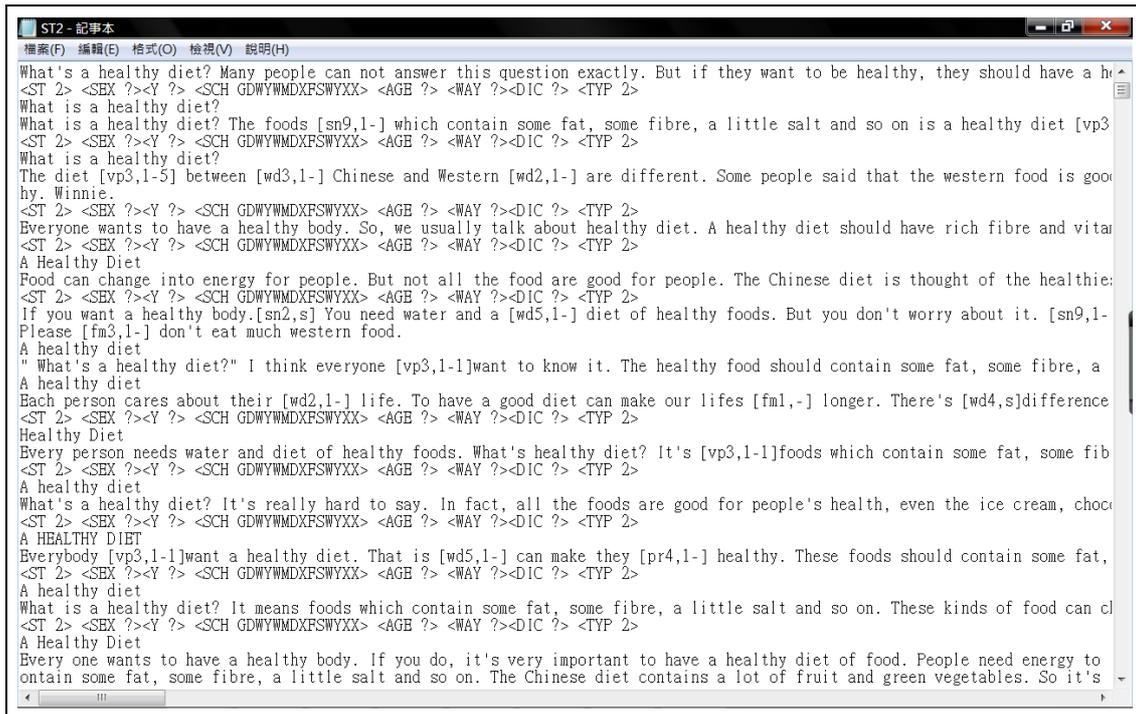
topics will also affect the results, but this information is not provided by the compilers in the profiles. It is unknown whether any one topic vastly outnumbered the others. Moreover, the topic effect will skew the results when the high-frequency PVs are concerned, as they have not been manually examined. This disadvantage can be amended by checking whether the collocates of the high frequency PVs match the topics. Bearing this in mind, I will interpret the results in the local context of the PV (within the sentence boundary), with the hope of reducing the topic effect as much as possible.

Guo (2006:50-55) elaborates the comparability issue between LOCNESS and COLEC. He acknowledges that both comparability and incomparability do exist between the two corpora. He compares the characteristics of these two corpora and points out that they are similar in aspects such as essay type, age of students, the authoritativeness of the compilers, and the time of completion (both were completed in 1998). They are also dissimilar in terms of the length of each essay, topics and genre. Guo argues that despite the small differences, the Chinese learner corpus and the reference corpus are comparable to a large extent. In addition, a completely comparable learner corpus and reference corpus cannot be found, since all of the variables would have to be controlled too tightly. In conclusion, such a comparison between CLEC and

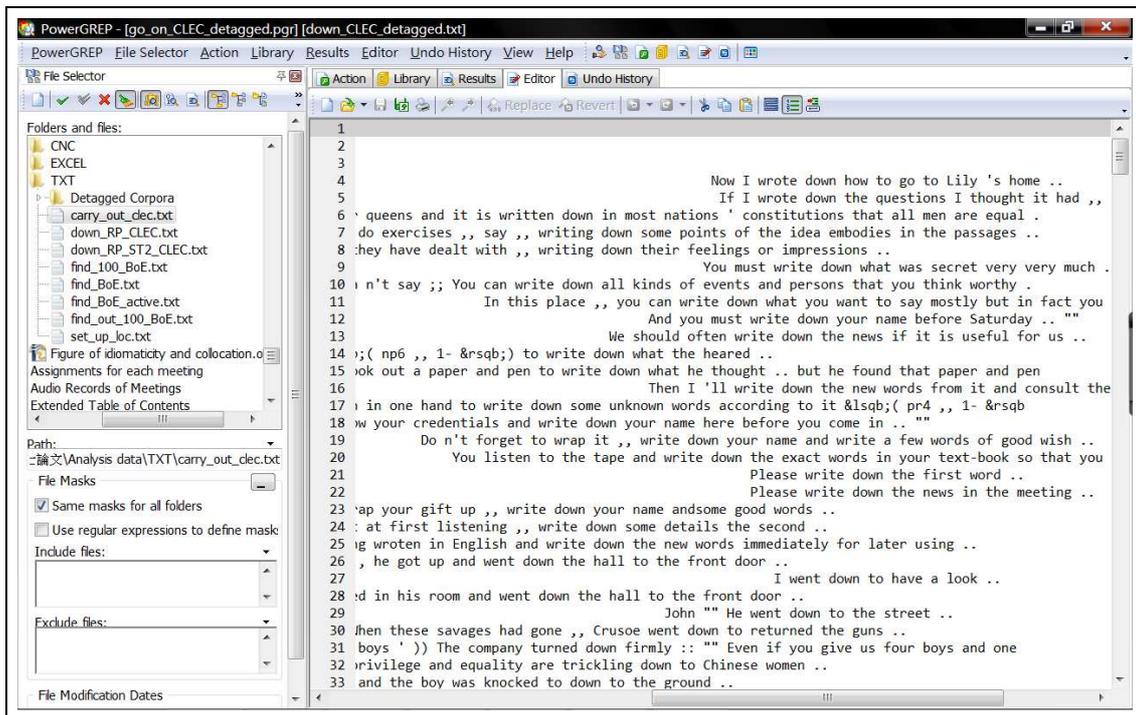
LOCNESS is not perfect, but is the best that can be performed in the present circumstances (see Guo, 2006:54-55).

### **5.2.2 Software**

The software package WordSmith4.0 is used to extract the data (Scott, 2004). The main program used is Concord, which is a concordance tool serving to sort the target items in their full original context. The concordancer is a basic but powerful tool which allows researchers to manipulate language data according to their needs (Hunston, 2002:38-66). Besides WordSmith, other software packages were employed where necessary. CLEC has been error-tagged and annotated with some additional information, but these error tags and annotations sometimes interfere with data processing; as a result, a computer program, PowerGREP 3.5 (Just Great Software 2008) was used to detag CLEC and clean the data (see Figure 5.1 and Figure 5.2).



**Figure 5.1: A screenshot of the original data with error tags and identity annotations**



**Figure 5.2: A screenshot of the data cleaned by PowerGREP**

### **5.2.3 Materials**

This thesis does not intend to examine all PVs with possible particles exhaustively, but aims to provide a thorough analysis of PVs with certain particles, which are anticipated to yield interesting results. Five particles were randomly selected for further investigation. They are UP, OUT, ON, ABOUT, and DOWN; the first two will be dealt with in Chapters 6 and 7 respectively, and the other three will be discussed in one chapter (Chapter 8).

## **5.3 Procedure**

### **5.3.1 Extraction of PVs**

This present thesis will deal with PVs using automatic and manual methods at the same time, in order to answer Ball's (1994:295) call for extracting data in both ways. PVs have drawn many researchers' attention in the field of natural language processing owing to the difficulty of extraction (Baldwin & Villavicencio, 2002). In computational studies, the targets are often termed VPCs (Verb-Particle Constructions) but not phrasal verbs, although they are referred to as similar constructions. Baldwin and Villavicencio (2002) point out that automatic extraction of VPCs is not easy, because the verb and the particle are not necessarily contiguous. Many difficulties will crop up and

simultaneously generate non-targets/noise when solely automatic extraction is adopted; these difficulties will be discussed in the following sections. In order to avoid complete reliance on automatic software which produces noise which may distort the results, a human manual check of the data is required. Therefore, a semi-automatic procedure that combines manual investigation with the help of computer programs will be employed.

#### *5.3.1.1 Criteria of selecting research targets*

A more important issue is the identification of target PVs, i.e. what constructions qualify for investigation. As seen in Chapter 2, there are many subtypes of verb-particle constructions, and researchers have used different conditions to define them. They consider different groups of these subtypes as PVs; some take a freer perspective and allow more subtypes and some tend to restrict the definition of PVs.

In the first place, the criteria deciding which groups of these verb-particle constructions are to be studied in this thesis have to be established. Deciding what constructions are to be included in our research is not simple. Two questions need to be considered. The first is the acceptance/rejection of constructions which contain a non-adverbial particle, including prepositional verbs and phrasal-prepositional verbs.

In Chapter 2 we have seen that prepositional verbs and phrasal verbs are mostly dissimilar in their characteristics apart from both being in the V+P form. Waibel (2007:63) argues that ‘prepositional verbs’ should not be taken into account when studying phrasal verbs, because learners face very different problems dealing with these two groups. For phrasal verbs, they have difficulties in perceiving the idiomaticity; for prepositional verbs, the problem lies in the correct selection of the prepositions. It seems better not to cover prepositional verbs in our analysis. However, we also witnessed that dividing prepositional verbs and phrasal verbs is not an easy task, since ambiguous instances can always be found (see Chapter 2), therefore I have had to resort to the use of POS-tagging (see the details of CLAWS below). As the extraction procedure proceeds, prepositional verbs will be left out of the analysis naturally, thus our discussion can focus on phrasal verbs which have an adverbial particle only.

Another group of constructions which may provoke disputes are the phrasal-prepositional verbs, i.e. constructions of three elements like *come up to*, *give over to*, *do away with*, etc. It can be argued that this group should be considered separately from PVs because these three-word strings have some common properties, such as always occurring in continuous sequence, and all the elements are compulsorily presented and fixed as three-element constructions, but not like two-word constructions

as PVs. However, such a separation of PVs and phrasal-prepositional verbs is meaningless if the phrasal-prepositional verbs are taken as extensions of PVs. Because our method is to extract the two-word V+P constructions first, and then study other co-occurring words in a span, the phrasal-prepositional verbs will be discussed as variants of usage as well, and are regarded as extended phraseological units of the PV they contained. The term ‘phrasal verbs’ in this thesis includes the traditional ‘phrasal verbs’ and ‘phrasal-prepositional verbs’, but excludes ‘prepositional verbs’.

The second issue concerns the idiomaticity of phrasal verbs. Those verb-particle combinations which are idiomatic or figurative are often accepted as phrasal verbs without much controversy, but others which are transparent or literal are liable to dispute. The complication of the idiomaticity of PVs has also been reviewed in Chapter 2. It is evident that problems will always arise if PVs are classified by their idiomaticity degrees. PVs can be evaluated by how opaque they are, but they cannot be put into groups, because the degrees are relative. Waibel (2007:63) points out that free combinations/literal PVs and idiomatic/figurative PVs cannot easily be separated, therefore she chooses not to divide them. I agree with this conclusion; thus such a differentiation will not be made in this present study. As a result, the free combinations (the particle is an obvious adverb which can be substituted by other adverbs easily and

the verb is often an action verb, e.g. *When I heard that sound I was afraid, and then I ran out*) will be included in my data, because firstly they are not easily separated from others, and secondly they constitute a considerable portion of the data, especially in the learner corpus. Similar criteria of identification have also been accepted by Liu (2011:663-664), who advocates that a simple syntactic criterion (i.e. a lexical verb with one adverbial particle) works better than an indirect and complex semantic one (i.e. new idiomatic meaning instead of the straightforward meaning of a verb and a particle). To conclude, the PVs examined in this research are those which have adverbial particles (thus excluding prepositions), and the issue of idiomaticity is not considered (thus including free combinations).

The criteria for the definition of PVs will certainly affect the quantitative results when the overall numbers of PVs are counted, as what was examined was based on the research question regarding tokens and types of PVs. Excluding some verb-particle combinations will not affect/impair the qualitative analyses, as these examine selected instances in detail. The criteria will only be applied to take out targets which are not suitable for detailed inspection.

### 5.3.1.2 *Cleaning non-targets*

After deciding the target items to be included, we come to the methods of extraction. First, the error tags which were annotated in the original CLEC have to be removed. This is achieved by running the software mentioned in Section 5.2.2. At this point, it seems any of the two-word constructions V+P in the corpora are ready to be culled by simply carrying out a search of the particles. For example, if we are to probe the Verb + OUT construction, firstly all of the combinations of Verb + OUT have to be singled out from CLEC and LOCNESS, in order to produce a frequency list of each verb type of the Verb + OUT construction. To retrieve all of the instances of the Verb + OUT construction, we can start by searching for the particle: querying OUT in the corpus. The computer program Concord will return all of the cases of OUT, which are messy and numerous. At this point, the duplicated instances were found and deleted with the assistance of WordSmith4.0. The results were skimmed to eliminate obvious non-targets. Unfortunately, the data retrieved in such an approach contains too many cases which do not meet our definition of PVs. Thousands of examples can be returned when one particular particle is searched. The numbers of the instances are too large to perform a detailed analysis; the worst part is most of the cases are not the true PVs we are looking for; therefore the data has to be distilled using a more precise procedure.

Clearly it is more effective to screen the data with the assistance of POS-tagging. A fully automatic method which uses POS tags is adopted by Gardner and Davies (2007:341); their corpus study of PVs uses a simple but functional definition of PVs. They search all two-word verbs which are tagged as a lexical verb and followed by an adverbial particle, either adjacent or not. They rely entirely on the validity of the corpus and tags, and no classification tests are done. A study which surveys a bulk of data must be extremely time-consuming, therefore an automatic approach is often adopted when the sheer number of distributional results (i.e. raw frequencies) is considered.

However, such an approach may save time and effort at the expense of data adequacy. The reliability of tagging PVs by an automatic tagger is doubted by Waibel (2007:67). The two-word constructions collected may not all be PVs in the narrowest sense: for example, the method could generate some accidental combinations of a verb followed by an adverbial particle. At the same time, the results may suffer from some data loss due to the error rates of the tagging system and computer programs. Moreover, some instances may also not be culled from the corpus, particularly when dealing with learner corpora, because presumably learners do not show as much consistency as native speakers, and the errors produced by the learners will often interfere with the

correct tagging. Therefore, the accuracy rate of the tagging may decrease and result in incorrect figures.

Despite these shortcomings, not using an automatic approach will demand unaffordable time and effort, rendering this research unachievable. The compromise is to adopt automatic and manual approaches at the points where either of them presents a clear advantage. It is hoped to adjust for the shortcomings of each approach by doing this. As I am fully aware that an automatic approach will have pitfalls, in order to decrease the effects of data inaccuracy and data loss, two measures are taken in this study. First, more precision is demanded for the qualitative analysis, therefore the data were not only filtered by the CLAWS annotation to get rid of the large amount of non-PVs, but each example of the filtered results was also manually and carefully probed. The problem that irrelevant non-PVs may not be completely removed can thus be solved, but the loss of data remains. Second, also in order to prevent the possibility of some unwanted instances still slipping through the tagging filter, when the individual cases of PVs are put forward for qualitative study later, a number of syntactic tests will be used to refine the data in the following phase, which will be presented later.

### 5.3.1.3 *The application of CLAWS*

According to our definition of PVs in this thesis, the particle has to be an adverb. To facilitate data capture, CLEC and LOCNESS were both POS-tagged by CLAWS, which is a tagging system developed by Lancaster University from early the 1980s. In the tagging guide of CLAWS, the tag RP is assigned to candidate constructions, which are termed as prepositional adverbs/particles. (This category is listed under both the sections of adverbs and prepositions.) In this tagging guideline, the author explains:

We assign the tag **RP** to a preposition-type word which has no complement. Typical uses of RP are in phrasal verb constructions, or when it functions as a place adjunct.

e.g.

*there's a lot of it <w RP> about these days*

*Don't give <w RP> up on us just yet.*

After this example, the author provides a full list of possible RP words: *bout, about, along, around, back, by, down, in, off, on, out, over, round, through, thru, to, under, up.*

The author also points out that the most crucial problem in assigning the tag RP correctly is the disambiguation of prepositions (tagged as II) and prepositional adverbs/particles (tagged as RP). The demonstration of disambiguation is given by the author through two examples (a) and (b) below:

(a) *She ran <w II>down the hill.*

(b) *She ran <w RP>down her best friends.*

In (a), *down* is a preposition, because:

(1) An adverb could be inserted before it: *She ran quickly down the hill.*

(But not: \**She ran viciously down her best friends.*)

(2) It can be moved (somewhat awkwardly) to the front of a *wh*-word:

*This is the hill <w II>down which he ran.*

*<w II>Down which slopes do you like ski-ing?*

In (b), *down* is an adverbial particle because:

(1) It can be placed before or after the noun phrase acting as the object of the verb:

*She ran her best friends <w RP>down.* (But not: \**She ran the hill down.*)

(2) If the noun phrase is replaced by a pronoun, the pronoun has to be placed in front of the particle:

*She ran them <w RP>down.* (= her best friends)

(But not: \**She ran down them.*)

Similarly: *The dentist took all my teeth <w RP>out.* ~ *The dentist took them out.*

Contrast: *She went <w II>through the gate.* ~ *She went through it.*

The above examples and explanation made by the CLAWS researchers clearly show the principles that work behind the scene. By using such a tagger, it is hoped that PVs can be extracted more efficiently. The POS tag RP separates the prepositional adverbs/particles from the general prepositions (tagged as \_II), thus the unwanted instances in our data (e.g. *keep\_VV0 pushing\_VVG it\_PPH1 up\_II the\_AT hill\_NN1...*) can be filtered out. The purpose of this step is to discriminate and discard the non-particles in the constructions such as those mentioned above.

How CLAWS works to automatically tag a corpus can be understood by the tagging process of BNC. The automatic tagging process of CLAWS runs through six stages: tokenisation, initial tag assignment, tag selection (disambiguation), idiom-tagging, template tagger and post-processing. The first stage, tokenisation, counts the word tokens and orthographic sentences separated by spaces and sentence boundaries. Then the second stage, initial tag assignment, assigns one or more tags to the words according to a reference lexicon and chooses the most probable tag. The next stage of disambiguation also adopts a probabilistic method, *Viterbi alignment*, to estimate the likelihoods of tag sequences, thus disambiguating confusions. Some special cases such as multi-words are better tagged as one unit, so some rules will be applied at the stage of idiom-tagging. An additional error-correcting piece of software, the template tagger, is

designed to supplement the insufficiency of the earlier stages. The final phase, post-processing, aims to provide ambiguity tags which allow the presence of two possible tags. Through these procedures, CLAWS is able to produce as accurate an output as possible (for details, see the BNC2 POS-tagging Manual online).

However, it is hard for any tagger to achieve a zero percent error rate, and so it is with CLAWS. It is claimed to have a 96-97% accuracy rate (see the website of CLAWS). Unfortunately, this accuracy rate is measured for common words: it is not clear how accurately CLAWS can deal with RP tags, especially in a learner corpus with errors. Take the particle ON for example: if we simply searched the word ON, 6504 instances were retrieved from the original data of CLEC, but when tagged with CLAWS, only 357 instances were retained. For the LOCNESS data, 1804 instances were found from the raw data, but the CLAWS tagged data returned only 152 instances (c.f. Chapter 8). Most of the instances eliminated from the non-tagged data are not V + ON constructions (i.e. PVs): in other words, in these instances, the particles function as prepositions but not adverbs. In order to estimate the accuracy tagging rate, 100 random instances were taken from the initial untreated corpus and the numbers of accurately tagged instances were counted. The result shows that approximately 90% of the two corpora are correct.

With the help of CLAWS, the program is believed to capture most of the probable candidates of phrasal verbs to its best ability. However, a small number of errors may still occur in the automatically-filtered data. For example, the two instances below both contain particles tagged as RP by CLAWS, but the particles do not form V+P constructions with verbs. This kind of instances are removed as there is no verb available.

[5-1] Then\_RT the\_AT elated\_JJ man\_NN1 march\_NN1 in\_II procession\_NN1 with\_IW nothing\_PN1 **on\_RP** .\_.

[5-2] From\_II then\_RT **on\_RP** ,\_, I\_PPIS1 became\_VVD like\_II to\_II by\_II air\_NN1 .\_.

Also sometimes a few general prepositions were not filtered out (see example [5-3] below), and moreover, the RP tags may sometimes contain prepositional verbs, as evidenced by the examples [5-4] and [5-5] below taken from LOCNESS which have the prepositional verb '*rely on*'.

[5-3] They\_PPHS2 **play\_VV0 on\_RP** the\_AT street\_NN1 with\_IW all\_DB kinds\_NN2 of\_IO colourful.

[5-4] I\_giving\_VVG others\_NN2 support\_VV0 ,\_, as\_II31 well\_II32 as\_II33 a\_AT1 shoulder\_NN1 to\_TO **rely\_VVI on\_RP** when\_CS feeling\_VVG weak\_JJ .

[5-5] What\_DDQ will\_VM our\_APPGE social\_NN1 development\_NN1 **rely\_VV0 on\_RP** if\_CS the\_AT market\_NN1 is\_VBZ full\_JJ of\_IO fake\_JJ commodities\_NN2 .

We can see the majority of prepositional verbs can be screened by CLAWS, but a few of them may still escape from the filtering procedure. These cases can be removed by applying the syntactic tests proposed by Darwin and Gray (1999:77-81) (see Section 2.5.2.3 for details of the tests), but I decided not to perform a comprehensive check on all the PVs. The reasons are twofold: firstly it is time-consuming to apply these five tests to each example of the PVs, and scrutinising these cases will put us off the track. Moreover, the multi-senses a PV have may severely aggravate the situation. Secondly, the numbers of occurrences are usually fairly small and will not significantly influence the quantitative results. In consequence, they will be kept in the frequency lists (Appendix A-C). However, the targets selected for detailed qualitative studies have to be real phrasal verbs, thus I will apply the tests to these targets. Therefore only those selected for further analysis will be examined by the five syntactic tests, in order to confirm their authenticity as true PVs.

Let us now turn to another problem: learners' errors. Their errors have two types: the first is grammatical errors, which have a mild influence on our analysis; the second involves learners' creation of illegal combinations. The recognisable errors were corrected and included in the list. For instance, *freshen up* will be collected as a use in

CLEC, although the original text produced by the Chinese learners is “*we will be fresh up*”. Other errors or misuses which have nothing to do with PVs were not identified at this point, but will be isolated later only if they affect our analysis. Those cases where a PV cannot be easily recognised were removed from the list.

So far, the data has been screened but the remainder still contains the second type of error, illegitimate combinations created by the Chinese learners. For example, the Chinese learners invented examples such as *affect on, wheel on, jump down, hit down*. Such examples have a value in revealing the Chinese learners' lexicon of PVs, thus are retained in the lists which present the numbers of each phrasal verb, but they will not be analysed in the qualitative comparisons.

Also note that these results will include literal uses which are superficially the same as idiomatic uses, e.g. the literal *go up* ‘something rises’ and the idiomatic *go up* ‘to suddenly explode’. According to our definitions of PVs, the literal uses will not be set apart but will be taken as targets for analysis.

The PVs are grouped by the five particles UP, OUT, ON, ABOUT, DOWN. Within each group, certain phrasal verbs are picked out for case study. The list of these PVs comprises: DRAW UP, LOOK UP, BRING UP, GROW UP, PICK UP, CARRY OUT,

FIND OUT, GO ON, TAKE ON, BRING ABOUT, COME ABOUT, BREAK DOWN, CUT DOWN. The analysis of these phrasal verbs will be reported in Chapters 6, 7 and 8.

### **5.3.2 Extraction of phraseological units: collocations, associations and sequences**

When it comes to the analysis of phraseological units, two points need to be taken into account. The first is the way to categorise PVs into meaningful groups in which one structure holds one corresponding relationship with one meaning. This mainly concerns the disambiguation of the multi-senses a PV may have. We have seen that one form of a PV can have more than one sense. In fact, most PVs are polysemous; as found by Gardner and Davies (2007); on average 5.6 senses can be discerned per PV among the PVs they surveyed. The nature of polysemy signifies the need to decipher it, giving the subtle clarifications of each meaning that a PV bears.

Note that Gardner and Davies (2007) argues that the conceptualisation of what a word is, such as its multi-word/collocational nature, will influence the validity of corpus-generated findings. In recent theorising on languages, opinion has shifted from looking at word senses as a set of individual senses (as in dictionary entries) towards a ‘contextual view’ (Wible, 2008:172) of word sense, which considers the word’s use in

various contexts. When in empirical research this contextual view was put into practice through the contextual approach, in which word senses can be accounted for by contextual features, so to speak, the meaning is established by contextualising the word use.

As the usages of these PVs are highly related to their meanings, it is justifiable to identify and classify PVs into groups based on their meanings. However, as an intricate semantic tagging of the corpora under investigation is not available at present, the meanings of the target PVs in this present thesis have to be determined by myself according to the context, which can be represented by the co-occurring words. Therefore one crucial aspect of the analysis is that a PV will be recognised by its contextual features, which indicate the meaning/sense but not the superficial form. These contextual features are manifested by various phraseological units which construct the usages. The actual usages of PVs will be profiled and analysed in terms of phraseological information such as collocations (i.e. lexical associations), grammatical associations and semantic associations.

The extraction of phraseological units can be implemented by either the use of corpus tools or the researcher's introspection. Different units require different approaches. Collocates can be identified by a straightforward method relying on

computer programs which capture these units by the constant recurrence of their fixed elements, revealing consistent patterns that can be noticed. For more abstract units like semantic fields/associations/sequences, more responsibility lies with the analyst, who is required to deduce the points of similarity among the instances.

### *5.3.2.1 Finding collocates*

As collocates can be identified either automatically or manually, this work uses both kinds of methods. The automatic method is employed where massive amount of data is processed by computer programmes (e.g. LookUp for BoE and WordSmith for CLEC and LOCNESS). The words which frequently co-occur with a PV within a span of four words, as following Sinclair (1991:33), are extracted by the computer program. An issue which needs to be mentioned at this point is the measure of collocation strength, because the software requires values of strength to determine collocates. Therefore some statistical approaches are outlined here. Generally there are four methods of measurement that serve to determine and select collocates: t-score, Mutual-Information (MI) score, Log-likelihood, and z-score. Only the first three approaches are relevant to this thesis: the programs used include LL Wizard (Rayson, 2010), which takes the Log-likelihood score, LookUp in CobuildDirect, which produces both the t-score and MI-score, and WordSmith (Scott, 2004), which generates an MI and MI3 score. The

following will provide a brief introduction to t-score and MI-score, since they are commonly used for determining collocations.

A concise review of t-score and MI score is shown in Table 5.4 below, summarised from Hunston (2002) and Hanl (2012). The practical application of t-score and MI-score to corpus has been clearly illustrated in Hunston (2002:70-75), where the differences between these two methods of measurement are highlighted. The formulae to calculate the two scores in the table are adopted from Hanl (2012). For more detail in relation to t-score and MI-score, please see Oakes (1998). Although the program LookUp provides both t-score and MI-score for the lists of collocates, t-score is adopted when the collocates are extracted from the BoE (see Chapter 6), for the reason that the focus is not on identifying special collocations (rare or technical) but on general combinations; it is the reliability that needs to be assured. Log-likelihood scores are adopted when the over-/under-representation of PVs are compared. For the results yielded by the WordSmith program, MI-scores are used.

The extraction approach of collocations from CLEC and LOCNESS is different from that used for BoE, because the program LookUp in BoE returns the top 50 collocates by their t-score. But for the two corpora, since LookUp is a program designed for use with BoE exclusively, WordSmith4 is used instead to do the work.

However, WordSmith4 does not provide the t-score (only z-score, MI3 and MI, LL) and also the numbers of instances are much fewer: collocates will be retrieved by its 'Concordance Program', with the minimum frequency and length set to 2 and 1, calculated by MI.

**Table 5.4: T-score vs. MI-score**

	<b>T-score</b>	<b>MI-score</b>
Definitions	Indicates the degrees of confidence	Gives the extent of effect that the node word has on other words Measures association strength (how strongly are they associated?)
Critical value of significance	2 or higher (Hunston, 2002:72)	3 or higher (Hunston, 2002:71)
Calculation	$t = \frac{\text{SampleMean} - \text{Independence}}{\sqrt{\frac{\text{SampleMean}}{\text{CorpusSize}}}}$  $\text{SampleMean} = \frac{\text{CollocationTotal}}{\text{CorpusSize}}, \text{ i.e. } \frac{10}{100000000} = 0,0000001$ <p>Independence is calculated as follows:</p> $\text{Independence} = \frac{\text{CollocateFreq}}{\text{CorpusSize}} \times \frac{\text{Node}}{\text{CorpusSize}}$	$MI(x; y) = \log_2 \frac{p(xy) \times N}{p(x) \times p(y)}$  <p>p(xy): probability of collocations p(x), p(y): probability of individual words N: corpus size</p>
Disadvantages	<ul style="list-style-type: none"> <li>• Based on normal distribution which is not how language is constructed (Hanl, 2012)</li> <li>• Can only be applied to two-word collocations (Hanl, 2012)</li> <li>• Cannot be compared across corpora (Hunston, 2002:73)</li> </ul>	<ul style="list-style-type: none"> <li>• Not reliable for rare collocations (Hanl, 2012)</li> <li>• Does not take into account the overall observed co-occurrences, whilst t-score does</li> </ul>

The manual method is used to deal with manageable data (e.g. identifying collocates of specific PVs). The collocates will be scrutinised by the researcher, because the numbers of concordance lines are not large. The above-mentioned statistical approach extracts collocates of all grammatical categories, and the collocates do not necessarily have any relationship to the PV in question. The manual approach focuses on collocates which form meaningful units with the PV. Therefore controlling the semantic roles or part-of-speech of the collocates, as Nesselhauf's (2005) study which focuses only on the VN structure (i.e. restricting the noun to follow the verb and to act as the object of the verb), can help to locate more useful collocates. In this present research, the focus will also be limited to the V+N pattern (or N+V (N as the subject) because the direction is not taken into account), and the identification of collocates of PVs will be conditioned by their semantic roles (agents, patients, and theme<sup>i</sup>) and syntactic roles (nouns) in order to ensure that they have a stronger relationship to the PV.

#### ***5.3.2.2 Finding extended phraseological units***

Collocates can be easily identified either automatically or manually; however, more abstract and variable targets such as the extended phraseological units of a PV,

including the combinations/patterns/associations of node words, cannot be captured as easily as collocates. Of course software which serves to identify the most frequent phraseologies is available. A corpus tool function such as 'Pattern' in WordSmith works well, whereby some of the prefabricated expressions in a corpus can be identified easily, whereas the major pitfall of such an automatic program is that it is less powerful in solving the problem of variations. Abstract phraseological units such as semantic fields and semantic sequences are better recognised by human analysts. Therefore, all of the concordance lines of these selected PVs will be manually inspected one by one, in order to discover abstract extensions such as semantic associations and sequences. We will see in Chapter 8 that most of the abstract phraseological units are not instantly recognised by the 'Pattern' program but are noticed by the researcher when a line-by-line observation is made.

Note that because the purposes of the sub-projects in the result chapters vary, there are differences in the methods adopted, which will be described and explained where necessary. Furthermore, a methodological inconsistency will also be noticeable through the analysis process: that is, not all of the selected PVs will be analysed by all of the phraseological associations. For instance, a PV may be analysed by its collocates only, but not colligations, semantic fields or sequences. This is because not all of the

phraseological associations can be found in the context of one PV; only some associations have prominent patterns which can be identified. Thereby the PVs will be discussed case by case at the phraseological levels where they have interesting behaviours.

A final issue relating to the analysis procedure is how the numbers of PVs are to be counted. The polysemy of PVs raises the question of whether the frequencies should be determined by their senses or forms. This problem has been discussed by Waibel (2007:75), who explains that divisions of senses may be a better approach when dealing with a small amount of data; in addition, the ambiguities of learner language may mystify the clarifications of the different senses of a PV. This present study will at first present the total frequencies of each PV when the distribution is concerned, thus the PVs will be recorded by their lemma forms. As further examination continues, they will be probed by the distinct senses expressed by the same lemma form, as the phraseological context is closely related to the senses.

## **5.4 Terminology**

Some of the terminology employed throughout this thesis need to be explained. The first point is about two similar terms: ‘usage’ and ‘use’. By ‘usage’, I mean the ways in which a linguistic item behaves at all lexical, grammatical, and discourse levels. I

reserve 'use' mainly for the instantiation of actual strings of texts produced by the writers, in order to contrast with 'usage', but 'use' can also refer to the presence of something (e.g. the use of phraseology). A second point concerns the relationship between 'usage', 'pattern' and 'semantic sequence'. Since 'usage' describes word behaviours, it inevitably points to repeated and consistent 'patterns'. As such, these two terms can be used interchangeably at times. Besides functioning as a common term, a technical 'pattern' is also adopted wherever the conventions of Pattern Grammar are followed. That is, grammatical representations such as VN, V + *that* are also called patterns. The idea of 'semantic sequence' has been established in Chapter 4, but I use it as an all-inclusive term in this present thesis. Therefore, a 'semantic sequence' can be comprised of lexical items (i.e. exact words), grammatical patterns (e.g. V + N), meaning elements (e.g. [evaluation]) and even concepts (e.g. [purpose]). Meaning elements can usually only be realised by lexis within fairly limited range, but concepts are often discerned by a long stretch of text in a sentence (I do not look beyond the boundary of sentences in this research study).

## **5.5 Summary**

This chapter has outlined the reasons for corpus selection, corpus software use and the determination of the research targets, which set up the framework on which this present

study is based. It has also discussed in detail the procedures of extracting the phraseological units for analysis. Dealing with different units requires different extraction approaches. We will move on to the results chapters. The next three chapters consist of descriptions of the findings with respect to the PVs of five different particles.

# Chapter6: PHRASAL VERBS WITH THE PARTICLE ‘UP’

## 6.1 Introduction

This study will first explore the group of phrasal verbs which are made with the particle UP in the two corpora I am using. Frequency is used as the benchmark to measure the state of linguistic items in the majority of corpus studies. Mainstream learner corpus studies also employ frequency comparison to identify learner-specific features. One of the general approaches dealing with data distribution is to look at over/under-uses, pointed out by Granger (1998b:13) as a powerful tool to highlight non-nativeness and inform language teaching. This approach is advantageous in manifesting the large differences between two corpora. Taking overuse or underuse to discriminate native and non-native corpora can be seen in studies of vocabulary (Ringbom, 1998), conjuncts (Altenberg & Tapper, 1998), phrasal verbs (Sjölhom, 1995) and many others (Aarts & Granger, 1998; Granger & Rayson, 1998; Lorenz, 1998; Milton, 1998). However, researchers such as Guo (2006:45) have expressed concern that the two concepts ‘over-/under-uses’ may be over-generalised. Guo explains that comparative studies by their nature will certainly lead to findings of discrepancies in different degrees. Arguing the dichotomy division of over-/under-use is simplistic, he proposes

to compare linguistic items in eight categories such as 'large vs. large frequency' or 'large vs. small frequency' (Guo, 2006:175-179). Although his analysis provides some description of and reasons to bolster the merits of such a fine-grained comparison approach, such a fine category scheme is not appropriate for this current study.

Instead, the binary division of over-/under-use is regarded as sufficient to recognise learner-specific features and will be adopted in the present study. This decision is made on the following grounds: first, the measure of over-/under-use utilized in this work is underpinned by statistical support, whereas Guo's categories of large vs. small (or small vs. large) comparisons are not. His categories are plain descriptions of quantity from a relative perspective and may lack objectivity. Second, the goal is simple: We just need to know the verb types that are used more or less often by the Chinese learners. Guo's eight categories are actually an extended and complicated version of the binary over-/under-use division.

For the purpose of this current study, over-/under-use thus will be used to contrast CLEC and LOCNESS. This current study will use the term over/under-representation, which is also mentioned by Kaszubski (1998:177), rather than over/under-use to refer to the same concept, because the object to be described is the language occurring in a specific corpus, or the language that might be used by the speakers in all circumstances.

It makes more sense to use ‘representation’ rather than ‘use’ when two corpora show different features but their contents are not fully comparable. Other means often used in corpus studies also include the type-token ratio (Meunier & Granger, 1998), which serves to demonstrate lexical diversity.

In the first part of this chapter (Section 6.2), the distribution of the Verb + UP construction will be presented. The verb types of all of the phrasal verbs identified in the corpora will be listed and their frequencies compared. The variations of the frequently used PVs between the native and the Chinese learner language will be contrasted. This section also includes analyses of type-token ratios, and the examination of over/under-representations will be conducted as well.

The second part of this chapter comprises a detailed analysis of five specific PVs selected from the data. The phraseological units which have close relationships with these five PVs will be examined in Section 6.3. The focus will be shifted to the issues of idiomaticity and restriction strength, with examples of the five analysed PVs. A bi-axis illustration based on idiomaticity and restriction strength will be proposed to disambiguate the confusion of these two issues (Section 6.4).

To summarise, this chapter will answer various questions with respect to the PVs with UP:

- In terms of distribution, what are the frequencies of the PVs with UP in CLEC and LOCNESS? What are the most frequent PV types? What PVs are over/under-represented in the two corpora? What is the type-token ratio?
- In terms of phraseological units, how do the uses of PVs with UP in both corpora differ?
- Regarding the relationship between idiomaticity and restriction strength, how can the PVs be represented by these two dimensions?

## **6.2 Verb types, TTR and over/under-representation**

The complete frequencies of the PVs with the particle UP are displayed in Appendix A, in their lemma forms. The second and fourth columns represent the absolute frequencies of each PV; the third and fifth columns show the normalised frequencies per million words. 1630 instances were found from CLEC and 363 from LOCNESS. In terms of verb types, 101 and 80 types were employed in CLEC and LOCNESS respectively.

Type-token ratios (TTRs) will be calculated for the PVs of each particle group. Although TTR is not a measure without problems (Granger & Wynne, 2000; Mollet, Wray, Fitzpatrick, Wray, & Wright, 2010), when used carefully, it is still one of the easy and intuitive procedures that are commonly adopted in learner language studies (for example, Cadierno, 2004). In this present investigation, TTR can allow us to evaluate the diversity of the particle groups across the native and non-native writer groups.

The TTR used in this study is a modified version of the standard TTR because not every individual word type and token will be analysed. The purpose is to compare the proportions of PV types (lemmatised) on the basis of PV tokens in each particle group between CLEC and LOCNESS. By doing so, we can find the average quantity of PV types per one hundred cases of PVs which are used by the Chinese learners and native speakers, respectively.

The TTRs were calculated by dividing the number of PV types with the number of PV tokens and then converting to percentages. The TTRs in the data of PVs with UP are 6.2% (the PV types (101) are divided by the PV tokens (1630) and multiplied by 100) for CLEC and 22% for LOCNESS. A higher percentage of TTR indicates more diversity; thus, the native speakers are shown to be capable of making use of more PV

types.

The results of the most frequently used phrasal verbs are documented in Table 6.1. Only those whose relative frequencies over 90 are picked, to eradicate low frequency PVs which rank high but do not occur frequently enough. As seen from the table, all of the top PVs are different in the two corpora except one, GIVE UP, which ranks 2nd and 3rd in CLEC and LOCNESS respectively. This PV is also found to be pronounced in the German and Italian learner corpora (sub-corpora from ICLE) reported by Waibel (2007:92), and she makes the interpretation that sometimes this could be a result of ‘topic sensitivity’. The question whether the frequent presence of GIVE UP is influenced by the article topics will not be pursued in this study, because some of the titles are not available in CLEC, rendering the analysis of topic effects unfeasible.

**Table 6.1: The top PVs in CLEC and LOCNESS**

	<i>CLEC</i>			<i>LOCNESS</i>		
	<i>Verb type</i>	<i>Asb.</i>	<i>Rel.</i>	<i>Verb type</i>	<i>Asb.</i>	<i>Rel.</i>
1	GET	161	150	BRING	38	117
2	GIVE	157	147	END	30	93
3	USE	115	107	GIVE	30	93
4	MAKE	106	99	GROW	30	93
5	TAKE	101	94			
6	SET	100	93			

Note: Asb.=Absolute frequencies(raw frequencies); Rel.=Normalised frequencies (per million words)

The comparison of high frequency PVs tells us which items are widespread in individual corpora; the device of ‘over-/under-representation’ can reveal which items are pronounced in one corpus if the other corpus is applied as the standard. The PVs which differ largely in the two corpora can be discerned by comparing the normalised frequencies. All of the verbs were sent to the Log-likelihood Ratio (LLR) test<sup>ii</sup>, performed by the online calculation tool provided by Lancaster University. The LLR was chosen as the means to analyse over-/under-representations on account of its well-established theoretical basis for corpus comparison (see Rayson & Garside, 2000) and the advantage that it can deal with the absence of data (i.e., when the frequency is zero), along with taking corpus size into account. Table 6.2 gives the top five over-represented and under-represented PVs and their Log-likelihood values. A higher significance of the difference is indicated by a higher value.

**Table 6.2: The top five over-/under-represented PVs in CLEC and LOCNESS**

<i>Verb type</i>	<i>CLEC</i>		<i>LOCNESS</i>		<i>over-/under-representations</i>	<i>LL value</i>
	<i>Asb.</i>	<i>Rel.</i>	<i>Asb.</i>	<i>Rel.</i>		
<i>USE</i>	115	107	0	0	+	60.86
<i>GET</i>	161	150	4	12	+	59.21
<i>RISE</i>	27	25	0	0	+	14.29
<i>TAKE</i>	101	94	12	37	+	11.97
<i>KEEP</i>	49	46	3	9	+	11.75
<i>BRING</i>	15	14	38	117	-	55.66
<i>END</i>	13	12	30	93	-	41.71
<i>BACK</i>	1	1	8	25	-	17.59
<i>RUN</i>	5	5	11	34	-	14.87
<i>OPEN</i>	3	3	7	22	-	9.79

Note: "+" means "over-represented" and "-" means "under-represented".

Three of the five over-represented PVs: USE UP, GET UP, TAKE UP, are also the most frequent items in CLEC. As a matter of fact, the six most frequently-used PVs mentioned earlier are all over-represented. Besides these three items, the LL values (in brackets) of GIVE UP (+5.92), MAKE UP (+1.33) and SET UP (+5.54) also suggest that they are over-represented. The other two over-represented items, RISE UP, and KEEP UP, do not have high frequencies in the Chinese learner corpus (RISE UP occurs only 27 times and KEEP UP 49 times); however, the gaps between the occurrences in

the two corpora are large (the over-/under-representation measures the disparities). In other words, although these two PVs do not occur very frequently in CLEC, LOCNESS contains far lower numbers of them from a relative perspective. Among the group of under-represented PVs, BRING UP and END UP, are also the two most frequent PVs in LOCNESS. The large numbers of occurrences (their normalised frequencies are 117 and 93) naturally render these two verbs under-represented in CLEC. The other three items: BACK UP, RUN UP and OPEN UP, rarely occurred in CLEC but were used in a fair quantity by the native students. Comparing the two groups of over-/under-representations, we can see the influence of genre types. Verbs such as GET, RISE, TAKE in the over-represented group are used to describe activities in daily life, while BRING, END, BACK seem to have more relation with arguments, for example: *bring up an issue, end up with a result, etc.*

It turns out the investigation of over-/under-representations does not offer much useful information for further analyses of PVs, because it is performed in a relative view so that the over-represented items may have low frequencies, which will render case studies rather difficult. Thus I will not examine the over-/under-represented PVs in other particle groups (OUT in Chapter 7 and ON, ABOUT, DOWN in Chapter 8).

### 6.3 Analyses of five example PVs

In this section, five example PVs will be explored to accomplish two purposes. First, these five PVs will be used to discern the Chinese learner language features. Second, they will be used as examples to visualise the interaction between different degrees of idiomaticity and collocation restriction.

Five PVs will be carefully analysed as examples, since a thorough study of every instance of all the PVs is unfeasible for this thesis. The criteria adopted to select these five PVs must be explained at this point. The most straightforward approach is to choose the targets from the most frequent PVs found in CLEC from Appendix A. (The Chinese learner language is the focus of this thesis, so the five examples will be based on the learner corpus rather than the native one.) However, although the five PVs need to be selected from CLEC, their collocations and semantics will be determined through search and retrieval from the large-scale BoE (LOCNESS is not used where the frequencies are rather low), which can return the most adequate and comprehensive results. Accordingly, the immense frequencies of these most frequent PVs returned from BoE, which are often the basic verbs of high frequency (for example, *make*, *take*), will impede human manual analysis, also the large numbers will increase the burden when the multi-senses of one PV are to be classified later, i.e. grouping collocates

according to their semantic fields. These restrictions lead me to refrain from probing the most frequent PVs but instead turn to another approach.

As picking our research targets from the most frequent PVs is problematic, an alternative method is to choose PVs which satisfy two conditions. First, they are not extremely frequent but still have sufficient<sup>iii</sup> frequency to yield convincing results make close analysis possible. Second, it is better for them to have variations of attributes, because it is assumed that more constitutional heterogeneity can prevent most of the example PVs from falling into similar positions on the diagram (Figure 6.1) to be created in Section 6.4.2. These five example PVs will be laid out on Figure 6.1, illustrating the categorisation of PVs of different degrees of properties (idiomaticity and restriction strength).

DRAW UP    LOOK UP    BRING UP    GROW UP    PICK UP

Among the PVs which meet the two conditions, the above five phrasal verbs are chosen because, first, their occurrences CLEC are not extremely rare. The raw frequencies of these five PVs in the CLEC are listed in brackets: DRAW UP (16), LOOK UP (50), BRING UP (15), GROW UP (88), PICK UP (50). Second, they have different attributes, as displayed in Table 6.3. The second row shows the numbers of senses of these PVs, which were taken from CPVD (McCarthy & Walter, 2006:29, 88,

146, 198, 226-228) (For the senses, see Appendix E). Based on the senses listed in the dictionary and the potential classification (e.g. literal or figurative, see the third to sixth rows in the same table), the degrees of idiomaticity can be approximately envisioned. At this point, we still have no idea of the degrees of collocation restriction of each PV, which will be determined at a later stage.

**Table 6.3: Attributes of the five phrasal verbs**

Attributes	DRAW UP	BRING UP	LOOK UP	PICK UP	GROW UP
polysemy	+5	+3	+3	+23	+2
idiomatic/opaque/not transparent	+	+	+	+	-
figurative	-	-	+	+	+
literal (fixed)	-	+	+	+	+
completive/aspectual-		+	-	-	+

Note: + means Yes, - means No. The figure indicates the number of senses identified in CPVD.

### **6.3.1 The Chinese learner language performance: the five example PVs**

The first step is to examine whether the most frequent collocates of these PVs are similar in the two corpora, which was retrieved by WordSmith4. Table 6.4 and Table 6.5 display the results, which give a rough impression that in the Chinese learner language, DRAW UP is more strongly associated with *laws*, LOOK UP with *dictionaries/words*, and PICK UP with *telephone*. While the Chinese learners seem to

cling to these basic strings, the native speakers demonstrate their ability to employ other usages of the PVs, e.g. BRING UP + *a question/issue*, PICK UP + *the scent*. However, the pitfall of reliance purely on computer programs is the cost of data loss. Some PVs do not have any collocates identified by the software, due to the small number of instances in our data and the fact that the program can only capture words of exactly the same word form. To obtain a more accurate analysis of these PVs, each instance was carefully examined by the researcher and the results are discussed respectively in the following subsections.

**Table 6.4: Collocation of the five selected PVs in CLEC**

DRAW UP		LOOK UP		BRING UP		GROW UP		PICK UP	
Col.	Fre.	Col.	Fre.	Col.	Fre.	Col.	Fre.	Col.	Fre.
<b>laws</b>	5	dictionary	11	--	--	television	3	(tele)phone	4
		words	9			girls	2	piece	3
		newspaper	4					knife	2
		sky	2					job	2

Note: -- = No noun collocates are returned by WordSmith4.

**Table 6.5: Collocation of the five selected PVs in LOCNESS**

DRAW UP		LOOK UP		BRING UP		GROW UP		PICK UP	
Col.	Fre.	Col.	Fre.	Col.	Fre.	Col.	Fre.	Col.	Fre.
--	--	--	--	children	3	--	--	scent	2
				question	2				
				issue	2				

Note: -- = No noun collocates are returned by WordSmith4.

### 6.3.1.1 DRAW UP

Two collocates, *law(s)* (75%) and *plan(s)* (19%), are consistently used by the Chinese learners, while for the NSs, the most common object is *constitution(s)* (33%). The percentages in the brackets are the proportions of one collocate out of all of the instances of that PV. The collocates, *plan(s)* and *constitution(s)*, are not identified by WordSmith, as they are absent in the above tables; however, they are no less significant as they occur in nearly one third of the data. Clearly, native speakers have a much wider range of lexical items which collocate with DRAW UP. Native students employ more varied nouns with this PV, such as *guideline(s)*, *proposal(s)*, *rule(s)*, etc., among which some more specific items like *pros-and-con(s)* and *reform(s)* are also present, as shown by the evidence of DRAW UP in both corpora below (examples [6-1]-[6-5] from CLEC, examples [6-6]-[6-10] from LOCNESS):

[6-1] Although there is still a long way for us to go to **draw up** laws for euthanasia because the current situation is not suitable [CLEC]

[6-2] What is the case in China, then? Time may not ripe for **drawing up** laws for euthanasia according to the report in Beijing Daily. [CLEC]

[6-3] So, if we **draw up** corresponding laws to guide and supervise euthanasia...[CLEC]

[6-4] Though time is not yet ripe for **drawing up** laws for euthanasia according to the report in Beijing Daily [CLEC]

[6-5] They should **draw up** laws to punish the producer of fake commodities and detect every product [CLEC]

[6-6] They met to **draw up** tough drug testing proposals aimed at stamping out the use of anabolic [LOC]

[6-7] In response to the events of '68, the Faure reforms were **drawn up**. [LOC]

[6-8] The company would have to **draw up** rules for a code of practice, which would be made known to all employees [LOC]

[6-9] When people kill each other most of the time they aren't sitting around **drawing up** prols and con's for murder raps, so the death penalty has no effect...[LOC]

[6-10] ...medical community examined the vareous techniques, dangers and benefits, and **drew up** a detailed set of guidelines for labs working in this field.[LOC]

It can be reasonably suspected that the repetition of *draw up laws* in CLEC is a result of the topic, which is corroborated by the recurrences of the word *euthanasia* in examples [6-1]-[6-5].

The collocate *constitution*, which is frequently linked with DRAW UP by natives, is also of interest, given that not many verbs can combine with it in the sense of drafting or preparing. So when the Chinese learners intend to express the same idea ‘draft the constitution’, which verb will they call up instead of DRAW UP? With close sense, the phrasal verb SET UP is found to be substitutable with DRAW UP in CLEC (see below), albeit it can still be argued to be slightly different (SET UP does not necessarily imply writing down the ideas). Whether this results from inter-linguistic difference (Chinese and English) or other non-linguistic factors (e.g. writers’ intentions) may worth studying in the future.

[6-11] Lastly, we should set up a constitution to protect fresh water.[CLEC]

### **6.3.1.2 LOOK UP**

The frequencies of LOOK UP in both corpora constitute a large contrast: there are 50 occurrences in CLEC, but zero in LOCNESS. In the Chinese learners’ data, 41% are literal uses (‘lift up one’s eyes or head’) and 45% are PVs in the sense of ‘finding a piece of information’. The majority of uses in the latter sense are represented by the formulaic sequence *look up words in a dictionary* (91%); other collocates are *book* and

*PC*, each occurring only once. I also looked at the verb collocates of *WORD* in *LOCNESS*, but none are found related to the sense of ‘finding a piece of information’. This shows that this phraseological sequence is more conspicuous to the Chinese learners, which may be an instruction-induced effect, since *LOOK UP+ X+ (in/from+ [resource])* is a common phrase taught to Chinese students.

### 6.3.1.3 *BRING UP*

The most obvious difference is that no case with the sense ‘start to talk about something’ is found in *CLEC*. In contrast, 58% of the citations in *LOCNESS* are used in this sense. It is not surprising that native students employ *BRING UP* an *issue, point, question, etc.* more, which is not observable in *CLEC*. The reason may result from *LOCNESS* containing mainly argumentative essays. Here are some examples from *LOCNESS*:

[6-12] of whether the Prince of Wales may rule as king when being a divorcee. This **brings up** the issue of him being the Head of the Church of England and so the ma

[6-13] This rush to finish judicial business **brings up** the issue of the finality of the death penalty. Even with appeals, it

[6-14] They constantly **bring up** the point that coal mining is a very dangerous job.

[6-15] This action by the author **brings up** the credibility of the author and their values.

Note that two different types of subject, human and non-human, can be classified in these examples. Example [6-14] has the human subject *they*, thus the PV means ‘introducing something to the conversation’. Others have a non-human subject (e.g. *this, rush, action*), therefore are used in a metaphorical way. The meaning of the PV is better described as ‘to draw attention to something’, and this usage seems to be constantly related to the collocate *issue(s)*. I consulted three reference dictionaries, (*the Cambridge Phrasal Verbs Dictionary* (2006), *the Macmillian Phrasal Verbs Plus* (2005) and *the Collins Cobuild Advanced Learners’ Dictionary* (2006)) and found that the non-human uses are not listed in any of them.

The rare presence of this broad sense ‘mention’ in the Chinese learners’ language also leads us to an interesting question: when the Chinese learners talk about these subjects, what verbs are used instead of BRING UP in similar sense? We can inspect the collocates of BRING UP, say *question(s), issue(s)*, to find out what alternatives are used by the Chinese learners. 392 and 73 cases of *question(s)* and *issue(s)* are returned respectively. I looked for the synonymous verbs in the sentences where the verbs can be replaced by BRING UP. Those verbs identified with *question(s)* are PUT and RAISE such as in *put the questions, raise the question*. As for *issue(s)*, the only verb found is

RAISE, as exemplified below. This reveals that this idiomatic expression BRING UP *an issue/question...* is not as familiar as PUT/RAISE to the Chinese learners.

[6-16] Legalized In China The question of euthanasia raises serious moral issues, since it implies that active measures will be taken to terminate human...[CLEC]

When it comes to the sense of ‘look after or educate’, the collocates are quite consistent in both languages. On the whole, the collocates are human, and can be classified into two categories: the child who was brought up or his parents (or other older people) who brought him up. This shows that although BRING UP is often regarded as a typical phrasal verb, or at least a figurative one, the Chinese learners are not confused by its usage at all. The opaqueness/idiomaticity does not bring on much perplexity. This may be due to the explicitness of the concept, which is universal in human cognition. The second reason may be the clear restriction of the collocation in English. Although there may be variations of the collocates, basically they can be put in the well-defined semantic sets of [children] or [family].

If we look at the noun collocates only, it might seem that the Chinese learners can handle this PV as well as NSs. However, if the scope is extended to the whole sentence and focuses on the functions of the linguistic items around the PV, a large difference is revealed. By functions I mean the information which is added to describe BRING UP, either pointing out the place (where), the purpose (why), how a person is brought up (how), or raised by whom (who). (See Table 6.6).

**Table 6.6: Functions of linguistic items around BRING UP**

Types	CLEC		LOCNESS	
	N	%	N	%
who	1	7	0	0
where	3	21	0	0
why	0	0	6	38
how	5	36	8	50
N.A.	5	36	2	13
Total	14	100	16	100

Note: N.A.= no modification

88% of the citations in LOCNESS and 64% in CLEC are modified by these functions. Apparently, the NSs insert more information around BRING UP to modify it, making the clause or sentence more complex. Looking into the modification types, it

appears that the native students tend to describe the bringing-up process, evidenced by the 50% presence of how-type functions. The Chinese learners also favour this how-type, but there is an equal chance that they will choose no functions. Furthermore, the NSs more constantly emphasise the purpose or result a child was brought up to become (NSs 38% vs. NNSs 0%), as seen in the why-type, which indicates purposes/results. Besides the functions, difference is demonstrated by several formal tools as exemplified below. The most frequent formal tools used by NSs are *to*-infinitive (43%) and prepositional phrases by NNSs (44%).

This result suggests that even the Chinese learners are able to use the restricted collocations correctly, they still need more knowledge about how the particular PV is used in a wider context. This knowledge include the appropriate ‘formal tools’ to be used and, more importantly, the groups of ‘concepts’ which are more likely to be bundled together. The attraction of related concepts to one or more linguistic items can be accounted for by the notion of ‘semantic sequences’ (see Section 4.3.3), which I will return to discuss in Chapter 9.

- Adverbs/ Adverbial phrases

[6-17] They do their best to feed their babies well, to **bring up** them [wd1,3-] well. [CLEC]

[6-18] ...thereby creating closely bonded families and well **brought up** children.[LOC]

- Prepositional phrases

[6-19] She was **brought up** at Gateshead...[CLEC]

[6-20] The children of Argos are **brought up** in this atmosphere of guilt...[LOC]

[6-21] The citizens of the town, condemn each other and **bring** their children **up** with a strong feeling of remorse and guilt...[LOC]

- To-infinitives

[6-22] Children are **brought up** to repent and what is more important is...[LOC]

[6-23] Even children are **brought up** to feel remorse and guilt for something...[LOC]

- Adjectives (usually with *it*)

[6-24] if he found [vp6,-s] it is to [wd3,-1] hard to **bring up** his children.[CLEC]

[6-25] they are aware of how much more stressful it is to **bring up** children later in life. [LOC]

#### 6.3.1.4 GROW UP

Most examples of GROW UP are literal uses, but a few of them are used figuratively,

such as:

[6-26] But Voltaire's *Candide* would be appropriately labelled as a bildungsroman or a **growing up** novel.[LOC]

[6-27] but we both know [vp6,4-] that the friend-ship [fm1,-] **grows** [vp6,3-3] **up** [wd5,4-2] with us, with the following [wd3,3-1] days.[CLEC]

The literal uses of GROW UP can be categorised into two groups, with emphasis on the progressive ([6-28] and [6-29]) or resultant ([6-30] and [6-31]) status of growing up (see below). Emphasis on the progress indicates the dynamic, continuous period of time; on the other hand, emphasis on the result suggests a static, further stage into which an entity develops. Therefore, examples [6-28], [6-29] and [6-30], [6-31] can be rephrased as *in the progress of growing up* and *became an adult/more mature*. The findings show that NNSs tend to use more resultant meaning (51%), but NSs favour progressive meanings (71%). A plausible reason for this divergence may come from the fact that the construal of the world differs in individual languages, which needs to be verified. The subtle distinction between progress- and result- connotations shows that even when the

Chinese learners make no grammatical errors in employing a PV, there will still be some non-error deviations, which cannot be discerned by traditional error analysis. Such a new area may change our views of learner language, and it calls for more attention.

[6-28] I saw Bobby bron [fm1,-]. I saw Bobby **grow up**. I saw Bobby died [vp5,1-].[CLEC]

[6-29] Many people say that this 'idol' talk sets a bad example for children when **growing up**. [LOC]

[6-30] In those days, I was eager to learn English. Then I **grew up**. I was studying [fm1,-] English and can sing English songs.[CLEC]

[6-31] As children get older and **grow up** males are accepted wearing soft colours, such as a light pink.[LOC]

The Chinese learners are also found to stick to a couple of formal means to distinguish between the progressive and resultant cases. In CLEC, the progressive sense is often marked with the word *as* (12%), and the resultant sense is labelled using *when* (37%), or *now/today* (12%). This is attested by the following selected examples (See citations [6-32]-[6-37]). Note that the marking of these words is not an absolute criterion; in other words, the occurrences of these words do not necessarily assign the correspondent sense to GROW UP. This is a tendency with a few exceptional cases.

Different marking is found in LOCNESS. In the NS data, *as* and *when* are used in similar percentages (both 13%) to indicate progressive GROW UP ([6-38] and [6-39]); no case is found marked by either *when* or *now/today* to convey the resultant meaning. The dividing line between *as* and *when/now/today* in the native language is not as clear as in the learner one.

[6-32] As we grew up, we can [vp9,2-1] feel knowledge is very important for every one.[CLEC]

[6-33] As he grew up, his desire of probing the mystery of the things became more and more...[CLEC]

[6-34] But when I grow [vp6,-] up. I know that there were [vp6,-] not only beautiful things..[CLEC]

[6-35] I wanted to be a teacher when I grow up.[CLEC]

[6-36] Now, I grow up and I know a student should get to know the world outside the campus. [CLEC]

[6-37] Now we have grown up. We are for [wd5,-] in love with somebody.[CLEC]

[6-38] As children grow up, they learn morals from their religious community.[LOC]

[6-39] Many people said that this 'idol' talk sets a bad example for children when growing up. [LOC]

### 6.3.1.5 *PICK UP*

The Chinese learner data displays a remarkable preference for using the literal PICK UP, namely meaning ‘to lift something by the hands’ (48%), as compared to only 23% in LOCNESS. This is not surprising, as the genre types in CLEC are not entirely academic. Some examples of this literal use are listed below.

[6-40] ..a hare crashed [vp4,-] into a tree and died. He ran up to [pp2,-] and **picked it up**. [CLEC]

[6-41] ...we began to **pick up** the stones and threw [vp5,7-1] it to the dustbin. [CLEC]

Apart from this, the other salient characteristic of the NNS data is the incorrect use of PICK UP. In these erroneous uses where the students seem to confuse PICK with PICK UP (in the sense of ‘choose’ in citations [6-42] and [6-43], ‘collect crops’ in example [6-44]), the meaning of the verb becomes more appropriate if replaced by PICK.

[6-42] As for my way [wd3,3-] , I will **pick up** some instructive ways. I will work hard at them [pr1,s-] . [CLEC]

[6-43] And when meeting by chance. [sn9,-] [fm3,-] We **pick up** some general and afe [fm1,-] topics like what happens to a friend recently... [CLEC]

[6-44] One day, I **picked up** a piece of fruit growing on campus and was caught. [CLEC]

In contrast to the Chinese learners' dependence on literal uses, the native students demonstrate their ability to cover more figurative or idiomatic senses of PICK UP. Examples such as *pick up the scent*, *the industry began to pick up*, *pick up the mad cow's disease*, etc. are presented in LOCNESS, but none of these usages appear in CLEC, implying that the Chinese learners need to be exposed to as many usages of a PV as possible.

So far, we have seen the idiosyncratic behaviours of five phrasal verbs and the fact that their usages in CLEC stand in stark contrast to those in LOCNESS. In the following section, we will turn to the attempt to handle PVs by incorporating their idiomaticity and collocation restriction degrees, using the five PVs as examples.

## **6.4 A visual illustration of PVs**

### **6.4.1 The ambiguity of idiomaticity and collocation restriction**

When the study of PVs and collocation is concerned, two notions, 'idiomaticity' and 'collocation restriction' are particularly crucial. Idiomaticity is a major characteristic of PVs on which the definitions of PVs are generally based (see Chapter 2). Collocation restriction is an issue often discussed in the study of collocation, especially when

restricted collocations are examined (c.f. Sections 3.4.1 and 4.2.3). In Chapter 2, we noticed an inconsistency in the literature where researchers employed different criteria to classify PVs. Sometimes these two notions are mixed and have caused ambiguity.

Therefore, it is intended in this sub-section to clarify this ambiguity, and use these two notions to develop a visual illustration which categorises PVs with the two defining criteria in a two-dimensional model. These two concepts will be explained first (Section 6.4.1.1), followed by an account of the factors which result in the ambiguity (Section 6.4.1.2). Issues relating to the illustration will be addressed in Section 6.4.2.

#### ***6.4.1.1 Idiomaticity and collocation restriction***

These two notions are worth briefly revisiting at this point. ‘Idiomaticity’ was introduced in Section 2.5.2.2, where the summary of previous studies revealed that it is generally regarded to have both narrow and broad meanings. In the area of phraseology, ‘idiomaticity’ can denote a similar meaning to ‘semantic opacity’ and ‘structural stability’ (Cowie, 1998); while in the area of PVs, ‘idiomaticity’ mainly refers to ‘semantic opacity’.

‘Collocation restriction strength’, that is, ‘structure stability’ as mentioned by Cowie, is used interchangeably with ‘commutability’ and ‘substitutability’ (Cowie & Mackin,

1993; Cowie, 1994), and has also been defined in various terms (e.g. ‘collocability’, ‘selectivity’, ‘variability’, and ‘combinability’; see the summary of Nesselhauf (2005:277). As suggested by these labels, this indicates the numbers or range of collocates a base word can take.

A brief review of collocation has also been presented earlier (c.f. Sections 3.4.1 and 4.2). Collocation generally simply refers to the co-occurrence of words/lexemes (e.g. Palmer, 1981; Quirk, Greenbaum, Leech, & Svartvik, 1985), to the ‘tendency’ in languages that words come together as a common term (Sinclair, 1966), or denotes a technical linguistic phenomenon describing the co-occurrences of lexical words where certain restrictions are at work (Hunston, 2002:68). Although these interpretations may focus on different aspects, they overlap to a large extent.

When it comes to collocations of the narrow/technical sense, i.e. the co-occurrences of lexical words, restriction is by no means the most essential issue to be concerned. Collocation range (Cowie, 2005:16) or collocability (Barkema, 1996), indicates the possible quantity of collocates that a base word can take. So how is collocability measured in the literature?

Let us turn to the classifications of collocations. Collocations can be grouped into different sub-categories by the status of each element. Cowie and Howarth (1996:83) breaks collocations into ‘invariable collocation’, ‘collocation with limited choice at one point’, ‘collocation with limited choice at two points’ and ‘overlapping collocations’. The ‘invariable collocation’ refers to a sequence in which none of the elements is replaceable, such as *foot the bill*. A ‘collocation with limited choice at one point’ is an example like *give/allow/permit access*, where only one element can be substituted by a limited set of collocates. Likewise, the ‘collocation with limited choice at two points’ is a combination where two parts can be changed, e.g. *get/have/receive a lesson/tuition/instruction*. With the ‘overlapping collocations’, the idea is more complex and can be best explained with an example. Verbs such as *convey* and *communicate* can both collocate with nouns like *point* and *view*, but *convey* can combine with other nouns like *regrets*, *condolence*, while *communicate* cannot. They overlap in some collocates but not all.

Collocations can also be classified by the numbers of collocates a base word may combine with. Howarth (1996:102) and Nesselhauf (2003:225-226, 2005:30) use descriptions of collocation amount (i.e. restriction strength) to differentiate collocations. Howarth (1996:102) explicitly phrases his definitions of levels of restrictedness with

expressions like ‘an open set’, ‘a small number/range’. A similar approach is also adopted by Nesselhauf (2005:30), where she divides VN collocations by the criteria of ‘a large group’ or ‘a small but well-delimitable semantic group’ of noun collocates. This implies that the number of collocates denotes the restriction strength. Therefore, numbers of collocates are rendered to be the primary factor that represents restriction strength, and the quantity of numbers often agrees inversely with the magnitude of restriction strength. In other words, if a base word takes fewer collocates, the collocation restriction strength is stronger and vice versa. A further suggestion is that using an exact number to determine the restriction strength is rather unfeasible. The restriction strength is better estimated by grouping with vague description.

#### ***6.4.1.2 The problem of ambiguity***

The problem of ambiguity is that these two concepts are easily confused and their interaction is often mistakenly assumed.

The first reason may be the overlap of idiomaticity and collocation restriction. Although these two concepts have different definitions, they suggest similar tendencies to a great extent. As seen earlier, researchers such as Cowie (1998) include the two

ideas ‘semantic opacity’ and ‘collocability’ together under the term ‘idiomaticity’, inevitably leading to confusion. Furthermore, such an overlap is also observable in studies where their correlation is tacitly implied by the categories of MWU types employed by the researchers. For example, Grant and Bauer (2004:43) produce a list of the semantic classification of idioms. Some researchers use ‘transparent to opaque’ (Moon, 1998; Yorio, 1980), and others adopt ‘open collocation to restricted collocation’ (Cowie & Mackin, 1993; Howarth, 1998), along the continuum of idiomaticity. Putting the terms regarding idiomaticity (transparent/opaque) in parallel with terms of collocation restriction (open/restricted) certainly hints that the two ideas are analogous.

Indeed, degree of idiomaticity is apt to correspond to strength of restriction in general. In other words, more idiomaticity usually agrees with more restriction of collocates, e.g. the most idiomatic combinations, i.e. the idioms, have extremely limited collocates (they are often unchangeable). It is commonly assumed at the outset that a more idiom-like collocation (more semantic opaque and specialised in meaning) usually imposes more restriction on the collocates, because intuitively, semantic opacity and specific meanings require only a small number of collocates. For the example of *put down the dog*, only animals are permitted to co-occur with *put down*,

but with *put down the pen* in the literal use, it seems that unlimited things are possible to be put down. However, this postulation is not completely valid.

This is first pointed out by Barkema (1996) and Hudson (1998). Howarth (1996:32) also cautions that:

*all semantically opaque composites (the most opaque being idioms) are to some extent collocationally restricted; indeed, there is some degree of correlation between the two characteristics [semantic transparency and commutability]. However, not all collocationally restricted composites are opaque.*

Later he continues to explicate that the scales of ‘idiomaticity’ and ‘restrictedness’ do not match each other (Howarth, 1996:101):

*[...] figurative meaning will not in most cases determine the restrictedness of a collocation ... This lack of match between the 'literal' and 'figurative' distinction and the dividing line between 'free' and 'restricted' collocations is to be expected: the presence of a figurative sense is a necessary but not a sufficient condition of restrictedness.*

The fact that idiomaticity does not accord with collocation restriction can be evidenced by examples found in my analysis (see Table 6.8 later). For instance, *pick up*

*someone/something* (to collect) and *pick up the phone*. *Pick up someone/something* is regarded as an idiomatic PV because the sense 'to collect' is opaque, but it has a very wide range of collocates: simply any person or anything can be picked up when it is used in the sense of 'collecting someone or something'. *Pick up the phone*, on the other hand, is a transparent PV, but it can be used literally 'to lift up the receiver' (see citation [6-45] below) or figuratively 'to make a call or answer the phone' (see citations [6-46] and [6-47] below). The former usage is similar to *pick up the book* but the latter usage is rather institutionalised and fixed, and can collocate with only one single noun (extremely restricted). An idiomatic (at least not compositional) PV may also have flexibility in selecting many nouns from a semantic set; for instance, *the cat was run over by a car/truck/vehicle*. The PV *run over* is idiomatic but the possible collocates are not limited to a very small number, although they are still constrained within a semantic set. In consequence, there is not necessarily a correlation between these two notions.

[6-45] The receiver dropped from her hand. She was kneeling on the floor, trying to **pick up her phone**.

[6-46] if he wants someone to talk to he knows my number and can **pick up the phone** any time.

[6-47] If you fucking **pick up the phone**, I will kick your ass.

Another cause of ambiguity comes from the fact that researchers suggest different criteria for delimiting free combinations, restricted collocations and idioms. Nesselhauf (2005:16) notes that Hausmann (1989:1010) separates free combinations and restricted collocations using collocation restriction but divides restricted collocations and idioms by idiomaticity. Another researcher, Aisenstadt (1979, 1981), applies both criteria to distinguish idioms and restricted collocations, but only commutability (collocation restriction) to separate free and restricted collocations. The incongruity further muddles idiomaticity and collocation restriction together.

In a nutshell, we ought to bear in mind that idiomaticity and restriction of collocation do not entirely correlate. They are neither mutually reinforcing, nor in a cause-and-effect relationship. They are two closely related but disparate concepts, both of which play fairly important roles in studying collocations. Therefore, for the purpose of emphasising these two notions in relation to PVs, I will attempt to produce a visual illustration to represent and elucidate their relationship, from a pedagogy-orientated view.

#### **6.4.2 Bi-axis illustration of PVs**

We have learned above that idiomaticity and collocation restriction are two independent criteria that indicate different dimensions of collocations. With the

purpose of applying idiomaticity and collocation restriction specifically to PVs, the following section is intended to create a demonstration that incorporates both of the concepts.

Idiomaticity, as has been introduced, refers to the semantic opacity/transparency of the whole collocation, and its degrees rely on the numbers of non-literal elements (Cowie, 1994). The more elements in a collocation have an opaque meaning, more idiomatic the collocation is. Because PVs are the main concern of this study, only the phrasal verb itself will be considered for idiomaticity, and not its noun collocates. That is, the degrees of idiomaticity in this study indicate the semantic transparency of the PVs instead of taking other elements into account. Two reasons lead to this decision. Firstly, the idiomaticity of PVs themselves is often deemed a vital factor that blocks the learning of PVs (see Chapter 2). Learners tend to run into difficulties when they come across idiomatic PVs, which is part of the concern of this thesis, which is aiming to provide pedagogical suggestions. Secondly, in most collocations, the PV is more likely to be idiomatic than the noun. It is rare to find cases of a literal PV with a non-literal (figurative or idiomatic) noun or a non-literal PV and a noun. Furthermore, pure idioms (both elements have lost their original meaning) such as *come down the pike*<sup>iv</sup> are

outside the concerns of this study, as they are extremely uncommon and will not particularly puzzle learners as long as their meaning is revealed.

Regarding the measurement of collocation restriction, I will merge the statistical/frequency-based approach with the phraseological approach (c.f. Chapter 4) to determine the restriction strength. Such an integrated approach is not new, since it has been adopted by researchers like Benson et al. (1986), Herbst (1996) and Nation (2001), as noted by Nesselhauf (2005:17). The approach I used to extract the frequent collocates of the PVs is a 'frequency-based' one (cf. Chapter 4), which considers 'probability' so that potential collocates over a threshold can be identified. The phraseological approach works when the V + N pattern is considered. It is used to determine the degree of collocation restriction strength.

Collocation restriction strength is primarily decided by the number of collocates, but three additional issues are particularly relevant to collocation restrictions: source types, synonymous collocates, and directions. These three have been discussed at length in Nesselhauf (2005:19, 27, 28-29). She states that two types of restriction source are distinguished by linguists: selectional restrictions (Fodor & Katz, 1964) and collocational restrictions (Cruse, 1986:107). The former is defined by semantic relations, and the latter by arbitrariness (Nesselhauf, 2005:19). Selectional restrictions

constrain the presence of a lexical element in a collocation by meeting certain semantic requirements. Collocational restrictions account for the conditions with which a collocation needs to comply in a language system. For example, *hire* co-occurs with a number of nouns (*staff, clerk, secretary, worker, etc.*) which belong to one semantic set, so the combinations are conceived to be constrained under ‘selectional restrictions’ (Fontenelle, 2005:192). Another example of selectional restriction is *kill*, which requires an animate object; on the other hand, *shrug one’s shoulders* is an instance of collocational restriction where *shrug* demands nothing else but *shoulders* as the object (Nesselhauf, 2005:19, 33). It can be concluded that the sources of restrictions arise from either the inherent semantic meanings or conventions which are arbitrary. Unfortunately, this division is not without problem because first, attributing the cause of collocation to either of the two source types is not always clear cut; second, the collocation source of either ‘semantics’ or ‘conventions’ seems to depend on our presumption of language. The first reason is acknowledged in Nesselhauf’s (2005:31, 227) statement that delimitation between arbitrary and semantically-motivated restriction is problematic at some points because it depends on the intricacy of sense description. I will add a second reason, that it is a consequence of different presuppositions as to how meaning is formed. Take as examples *drink + water* and

*purse + lips/mouth*, which are listed in the categories of free combination (semantically motivated, selectional restriction) and restricted collocation (arbitrary, collocational restriction) respectively in Nesselhauf (2005:30). If we say that *water* is chosen because *drink* means ‘take liquid into body’, this also suggests that we accept that *drink* carries an inherent meaning that demands liquid as the object, thus the base word selects its collocates. In contrast, if we assume that the meaning of the base word is defined by its accompanying collocates, the meaning of *drink* being restricted to ‘take liquid’ is inferred from the conventional collocations *drink + water/wine/coke*, as with the instance of *hire + staff/clerk/worker*. The same applies to the second example, *purse + lips/mouth*, which can also be regarded to be semantically motivated, given that *purse* can also be defined to have the meaning ‘contract one’s lips into a rounded shape’, thus the verb can only select *lips/mouth* to combine.

Nesselhauf (2005:27) also points out another problem related to collocation: the synonym issue, whether the restriction strength should be decided by synonymous collocates, i.e. collocates belong to one semantic set. This problem has not been explicitly addressed in the literature, but can be revealed from the examples given by different researchers, where an inconsistency can be discerned. The condition of collocates in one sense is clearly expressed by Howarth (1996:102) in his definitions of

restriction levels. An opposing view, however, is held by Aisenstadt (1979:73, 1981:55-56): in her examples of *shrug/square/hunch* + *shoulder*, the verb collocates do not have to be synonymous. Nesselhauf (2005:30) uses the criterion of semantic similarity as a means to describe the restriction extent. In her grouping of VN collocations, three categories involve some kinds of semantic restriction (e.g. *kill* + [+alive], *read* + [written material], *commit* + [something wrong]).

The next issue discussed is the direction of collocation. This is remarkable in cases such as verb-noun collocations. Nesselhauf (2005:42) gives the example that *commit* can collocate with a few words such as *suicide/crime/sin*, but *suicide* is only permitted to co-occur with *commit* and not other verbs. As a result, if *commit* is selected to be the base word, then the restriction strength of the collocation *commit* + [something wrong or illegal] is weaker than the collocation *commit suicide* when *suicide* is the base word.

In my study, the source types (selectional or collocational restrictions) will not be tackled, because such a distinction is on many occasions ambiguous, as has been argued earlier; therefore will not be considered in this study. The collocates will be grouped by the similarity of their senses, because in the two-dimensional model which I am going to present, the PVs will be probed by their individual meanings, so it makes more sense to discuss collocations within each semantic field. As to restriction direction, the

verb rather than the noun, as has been concluded by Nesselhauf (2005:29) is ‘semantically autonomous’ and should be assigned as the base word. The restriction direction in this study is, of course, from the PV to the noun collocates, as the verb is the base. Moreover, in order to make the collocations comparable, only the V + N (and N + V) pattern, in other words the combinations of one PV and the nouns which either operate as the agents or patients/themes to the PV, will be considered.

#### ***6.4.2.1 Finding collocates of the five PVs in BoE***

The typical collocations of the five selected PVs which have been analysed in Section 6.3 need to be identified in general English at this point, in order to measure their collocability. The frequent collocations of the five PVs will be investigated in BoE, which returns the top 50 collocates. Only the noun collocates (excluding pronouns) will be considered, because the focus of research is on the V + N (or N + V)<sup>v</sup> pattern. The selection procedure is exemplified with the phrasal verb GROW UP, as shown in Table 6.7, where only the top twenty collocates of GROW UP are displayed as examples for the sake of limited space. The noun collocates which are potentially the subjects or objects of the verb are emboldened and collected as our data. In this case of GROW UP, *children/kids/child/generation/boy* form collocations with the PV

**Table 6.7: Selecting targets among the top20 collocates of GROW UP in BoE  
(Span 4:4)**

<b>Collocates</b>	<b>Frequency</b>	<b>t-score</b>
i	3991	43.349608
in	6446	42.743817
<b>children</b>	1322	34.257322
when	1544	29.918438
who	1548	28.902382
she	1225	23.970294
with	2177	23.555752
where	769	22.730306
he	2084	22.515954
they	1492	21.086203
had	1092	16.874095
<b>kids</b>	304	16.462362
my	665	16.270901
<b>child</b>	315	15.699931
<b>generation</b>	219	14.137676
young	300	13.918996
ve	383	13.758123
you	1312	13.246848
<b>boy</b>	205	12.911694
born	204	12.894798

The results for the five selected phrasal verbs are listed in Table 6.8. The nominal collocates are first classified into groups in terms of broad semantic sets/fields (emboldened), and then the senses are identified based on CPVD (2006) (in capitals). It seems that some nouns are more likely to occur in the ‘subject’ position (in italics), whilst most of the collocates tend to be the ‘objects’ of the PV. The collocates’ being

subjects or objects is only suggested tentatively here, because they have not been tested in context. A few of the nouns which cannot be categorised into any of the semantic sets were grouped together under the category ‘others’. Words in this kind of group tend to be non-subjects/objects of the PV; for example, *action* and *contingency* are mainly the pre-modifier as in *action plan* or *contingency plan*. Some other examples of these collocates which are not directly related to the PV are shown in context below, where *way* and *part* also do not function as the roles of subjects or objects in the two citations.

[6-48] I think it's good to be brought up that way, because

[6-49] I want to see them grow up in this part of Europe

Two kinds of problem arose when grouping the collocates into semantic sets: labelling and categorising. Finding umbrella labels is not a simple task, as the collocates may be defined from different angles. For example, the collocates *child(ren)*, *kid(s)* under the superordinate term ‘young people/offspring’ can also be put under the labels ‘people’ or ‘human’. Naming the labels is quite difficult, as there is always the possibility of giving more specific meanings to the nouns. Categorisation of the collocates is also not as clear as is suggested by intuition because no absolute criterion can be relied on. The semantic sets are established on the semantic similarity they share, which is of some vagueness due to the inevitable subjectivity of the analyst judging it.

Besides labelling and categorising, one more insuperable problem is that a corpus cannot yield all collocate candidates of one semantic field, which will result in incomplete senses or inexhaustive collocates within one sense group. For instance, a search of DRAW UP in BoE retrieved forty-nine collocates (the last collocate is *legal* with a frequency of twenty-nine and t-score around 4.6), out of which one sense of DRAW UP ‘to stand up straight’ as in *He drew himself up when he talked to his superior* is difficult to recognise, because no collocate related to this sense is found among the frequent collocates. This makes sense, because DRAW UP in this sense is most likely associated with pronouns, people’s names or group labels, which are varied, thus no common word can be captured. Besides the absence of certain senses, it is difficult, if not almost impossible, to collect all of the acceptable collocates exhaustively. A massive corpus like BoE can capture the majority of collocates well, yet there is always the possibility of missing particular collocates owing to their extremely rare occurrences. For example, DRAW UP + *note(s)* seems to be a meaningful collocation, but only occurs five times in BoE. Since listing all possible collocates exhaustively is rather unlikely, I will only consider the top 50 collocates retrievable in BoE. From Table 6.8, the restriction strength of each phrasal verb can thus be determined.

**Table 6.8: Typical collocates of five example PVs**

<b>DRAW UP</b>	<b>LOOK UP</b>	<b>BRING UP</b>	<b>GROW UP</b>	<b>PICK UP</b>
TO PREPARE SOMETHING BY WRITING IT	LOOK UPWARD	LOOK AFTER	BECOME ADULT	ANSWER THE PHONE
<b>A collection of ideas</b>	<b>something above</b>	<b>young people/ offspring</b>	<b>young people/offspring</b>	<b>phone</b>
plans/plan	things	child/	child/children	phone/ telephone
list/lists	sky	children	kids	receiver
shortlist	ceiling	baby	boy/boys	
guidelines	stars	kids	girl/girls	
programme	face	sons	son	LIFT SOME-THING
proposals	eyes	daughter	generation	<b>anything</b>
draft	ADMIRE/	generation	<b>people</b>	ball
report	EMULATE	<b>family</b>	people	things
blueprint	someone	<i>family</i>		book
document	man	<i>mother</i>	<b>family</b>	bag
budget		<i>mom</i>	family	
strategy	FIND INFOR-MATION	<i>parents</i>	parents	LEARN
sheet	word		father	INFOR-MATION
	dictionary	START TO TALK ABOUT	mother	<b>ideas</b>
<b>Statements (enforceable by law)</b>		SOME-THING		points
	OTHERS	subject	OTHERS	tips

DRAW UP	LOOK UP	BRING UP	GROW UP	PICK UP
constitution	<u>desk</u>	issue	<u>part</u>	GO FASTER
agreement			<u>farm</u>	<b>speed</b>
code		FUNCTION AT A		speed
contract/contracts		DESIRED RATE/		pace
rules		HAVE THE LATEST INFO		
policy		speed		RECEIVE SIGNALS
legislation				signals
		OTHERS		
<b>People (with official power)</b>		<u>way</u>		PAY MONEY
<i>commission</i>				tab
<i>committee</i>				bill
<i>government</i>				
<i>officials</i>				COLLECT SOMEONE
<i>ministers</i>				car
<i>solicitor</i>				
TO MOVE A PIECE OF				OTHERS
FURNITURE				pieces
chair				injury

**DRAW UP**

**LOOK UP**

**BRING UP**

**GROW UP**

**PICK UP**

---

TO MOVE KNEES OR LEGS

CLOSER

knees

A VEHICLE STOPS AT

SOMEWHERE

car

OTHERS

action

contingency

task

---

#### ***6.4.2.2 Categorising PVs along the two axes***

As it is intended to present PVs on the basis of the two notions, degrees of idiomaticity and collocation restriction, by drawing on the five selected PVs, it is necessary to give an explanation of how the degrees of the two notions are divided.

The two notions are used as the two axes, and are termed ‘semantic transparency’ and ‘collocation restriction’ respectively to avoid confusion and misinterpretation. Along the axis representing degrees of ‘semantic transparency’, the three sub-categories are labelled ‘transparent/literal’, ‘semi-transparent’ and ‘opaque/idiomatic’ according to the semantic status of the PV. The ‘figurative’ use in terms of ‘extension from the literal sense’ is included in the ‘semi-transparent’ group, which also subsumes the completive/aspectual PV. Degrees of semantic transparency are decided by the meaning of the PV: whether it is literal, figurative/semi-transparent or idiomatic. The definitions of these terms have been elaborated in Chapter 2. A brief summary is shown below:

- Literal PVs: those whose meanings are the consequences of combining the inherent semantics of the verb and the particle. The verb usually maintains its original status and the particle either denotes directions (thus functions as an adverb) or simply adds no new meaning.
- Semi-transparent (figurative) PVs: those which have a metaphorical meaning derived from the original sense. This category also includes completive/aspectual PVs.
- Idiomatic PVs: the meaning is not available from either the verb or the particle, or the combination of them.

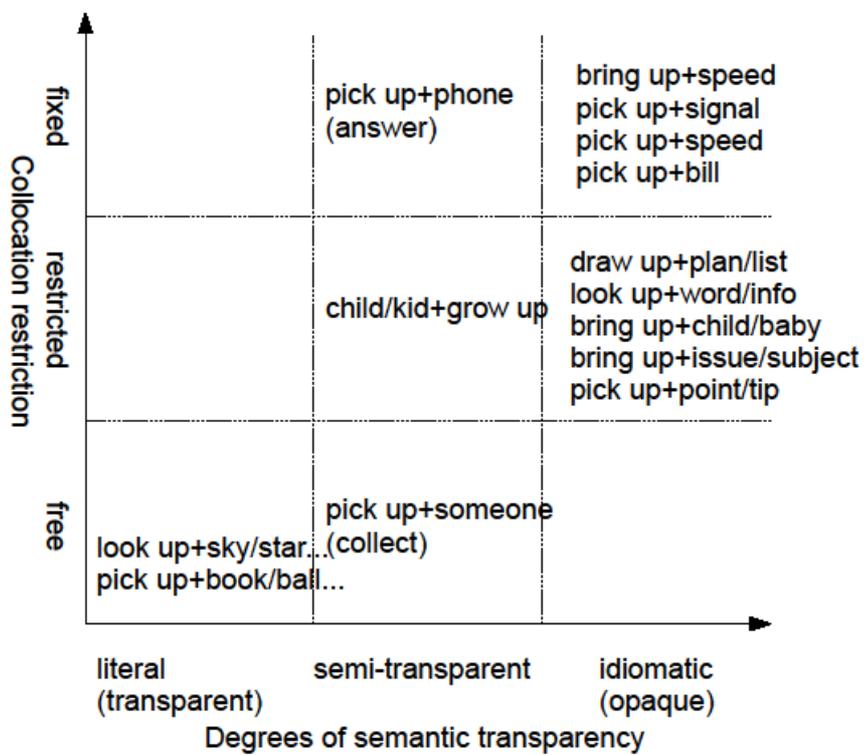
The other axis is the degree of collocation restriction, which is divided into ‘free’, ‘restricted’ and ‘fixed’. The strength of collocation restriction is judged by the quantity of collocates that a PV takes, which can be estimated from Table 6.8. The definitions of the three restriction levels adopted in this present study are:

- Free: the possible collocates appear to be unlimited; the only restriction condition is the semantics of the PV, which is usually general (e.g. PICK UP + *ball/book/toy/pen* etc.)

- Restricted: only certain collocates, which often can be delimited in terms of semantic fields, can be combined with the PV (e.g. DRAW UP + [ideas in the written form]: *plan/list*)
- Fixed: the collocates are limited to few nouns in a given sense (e.g. BRING UP + *speed*)

Placing the PV collocations in the appropriate positions can be exemplified by the representative instances of individual groups. The literal-and-free example is LOOK UP + *sky/sun/cloud/ceiling/star* (and their plurals), etc. in which the PV is transparent and the noun collocates seems to be unlimited. The idiomatic/opaque-and-fixed group is also easily discerned, e.g. BRING UP + *speed* is a typical combination that delivers a particular meaning. More than one collocate can be expected in the idiomatic/opaque-and-restricted group, whose noun elements can usually be described by a semantic field, e.g. DRAW UP + [ideas]. If the noun elements constitute a semantic set while the PV is figurative/semi-transparent, the collocation will fall into the middle category: semi-transparent-and-restricted, e.g. [young people] + GROW UP. Figurative uses of PICK UP in the senses of ‘answer’ and ‘collect’ are divided into two PVs respectively in the fixed and free categories, one with only one collocate *phone* and

the other with unlimited collocates. Other representative examples are also located in the illustration below.



**Figure 6.1: The two-dimensional model of PVs**

Undoubtedly, the classification of these PVs by their collocation restrictedness and degrees of idiomaticity is more or less subjective. In other words, the boundaries of the

two dimensions are not delimited with absolute criteria, so views of how the PVs may be allocated to the categories may differ. Cases such as PICK UP ('collect') and GROW UP are dubious. Is PICK UP in the sense of 'collect' an idiomatic PV or a figurative use? On the one hand, it is institutionalised and lexicalised that the verb and the particle are used as a whole unit (evidenced by the substitution of one word COLLECT), both giving up their original meanings to a certain degree. On the other hand, we can state that it derives from the literal use PICK UP 'lift something', and the meaning is extended from this base meaning to a figurative use, 'collect people/things'. The case of GROW UP displays a different situation. GROW UP can be seen as a literal PV which is equivalent to the simplex verb GROW, where the particle does not supply any additional meaning. Alternatively it can be taken as semi-transparent in the case of an aspectual PV, where the particle UP conveys the meaning of completeness, or in the case of a figurative PV with the meaning 'increment' carried by the particle UP. That said, the indeterminacy does not invalidate the bi-axis representation. After all, my intention in this illustration is not to give a hard-and-fast categorisation of PVs, but to illustrate that a PV can be represented by different degrees of semantic transparency and restriction strength of collocation at the same time. By incorporating the two aspects, PVs can be perceived and understood by learners with more efficiency.

Besides the representative PVs in the illustration, it might also be interesting to consider possible PVs which can be located in the two empty grids: the one that is transparent but fixed with one collocate, and the one that is opaque but free to co-occur with many collocates. Although none of our example PVs fit into these two categories, possible instances are conceivable. An instance for the former may be GO DOWN, which always collocates with *computer(s)* when it means ‘stop working’. The latter can be exemplified by RUB + OFF + (ON), which is idiomatic to learners but can be used to lead words of various kinds of people (*lads, players, me, you*) and others such as *golf* and *game*. These two examples are attested in BoE; however, PVs of these two categories are rare, and most cases scatter in the grids in between.

I only looked at the PVs in the UP group but not PVs with other particles in the visual illustration. The reason for this is that my ultimate goal in this section aims to develop a diagram that relates idiomaticity and collocation strength. It is not important to cover all of the phrasal verb groups with different particles. The comparison of PVs in the same group is sufficient to illustrate the idea, and similar results are expected when it is extended to groups of other particles. It is possible that one will find other PVs to be located differently on the diagram of idiomaticity and collocation strength.

Note that I do not claim that these five PVs represent the whole picture of all PVs, but only provide a preliminary example that can lead to future studies.

To sum up briefly, it has been verified that the two notions, semantic opacity and restriction of collocates, are not necessarily correlated, although they generally overlap. This leads to the idea of describing PVs using these two dimensions. The degrees of idiomaticity and collocation restriction, though being different notions, can converge to reach a coherent account of the complexities of PVs. It is hoped that representing the properties of PVs by means of ‘semantic transparency’ and ‘collocation restriction’ can help to account for the difficulties faced by learners.

## **6.5 Summary**

In the first half of this chapter, I have probed the UP phrasal verbs in the native and non-native corpora, in respect to their frequencies and their individual behaviours. The frequencies set up the foundation on which the sample PVs are selected. The results of the frequently used PVs and the discrepancies of type-token ratios and over-/under-representations disclose the phrasal verb items that require more attention and prove that the Chinese learner language, which has a low TTR, is less varied.

Five sample PVs, each thoroughly examined, reveal some significant findings. The first is that their prominent noun collocates in the Chinese learner language are different from those in the native language. Of course, this may be largely caused by topic or genre effects, but the fact that none of the collocates are the same may well suggest that the readily accessible nouns associated with a PV for the Chinese learners and natives are dissimilar. The second finding is that each PV exhibits unique behaviours, such that they cannot be studied using the same parameters, i.e. which linguistic phenomena to look for. Certain linguistic phenomena can be discerned in certain PVs but not in others. Above all, the parts of our analysis that taking more contextual elements (e.g. functions or semantic sequences) into account have proved fruitful in obtaining new discoveries.

In the second half, a diagram has been created that illustrates PVs on the basis of idiomaticity degrees and restriction strength, helping us better understand the properties of each phrasal verb. The layout of PVs by explicit separation of the two notions may have pedagogical value, in that teachers can introduce PVs of similar properties in an appropriate order, and the students can see the contrast of PVs having different properties rather than being only dimly aware of this phenomenon.

The results throw light on the fact that more features in depth are required to gain access to the whole performance of a PV, if we wish to see an accurate picture. This current chapter deals with the UP group, with the aim of investigating more PVs for fewer details; however, for the following chapters which examine other particles, fewer PVs will be studied but more details probed and the focus will be shifted to the phraseologies in terms of individual PVs. In the next chapter, the findings for the OUT phrasal verbs will be presented, with more emphasis on the surrounding context of the node PV.

# Chapter7: PHRASAL VERBS WITH THE PARTICLE ‘OUT’

## 7.1 Introduction

In the previous chapter, we looked at possible ways to pin down characteristics of the Chinese learners which differentiate them from native speakers and the relationship between ‘idiomaticity’ and ‘collocation restriction’. The results have informed us that analysing fewer examples in more depth may be a better way to study PVs. Furthermore, the evidence of one specific group, Verb + UP, may not be strong and valid enough. In this chapter, I will demonstrate that a contextual approach works effectively to disentangle the Chinese learners’ concealed differences by examining another group, Verb + OUT, with more focus on the phraseological behaviours of the verbs in this group. The questions to be answered are:

- In terms of distribution, what are the frequencies of PVs with OUT?  
What are the most frequent PV types in CLEC and LOCNESS? What is the type-token ratio?
- In terms of phraseological units, how do the uses of PVs with OUT in both corpora differ?
- What usage patterns distinguish two near-synonyms (one PV and one

single-word verb)? What problems do the Chinese learners have in using them?

Two phrasal verbs, CARRY OUT and FIND OUT, will be drawn on to illustrate how learner characteristics can be successfully identified. In Section 7.2, the results of one type of phrasal verbs, Verb + OUT, are presented, and a frequency list of all PVs with OUT in the two corpora is generated. Section 7.3 explores the phrasal verb CARRY OUT by looking at contextual features. Section 7.4 continues to probe another phrasal verb, FIND OUT, with an extensive study comparing FIND OUT with FIND. This chapter is concluded by a summary of the results.

## **7.2 Overall results**

The same procedure of data extraction as that described in Chapter 5 was employed to run through the corpora for the group of Verb + OUT. 1603 instances were found in CLEC and 434 in LOCNESS after filtering out the noise. All of the verbs which collocate with OUT are listed in Appendix B in alphabetical order. 142 verb types were used by the Chinese learners, and 108 types by the native students. It appears that more verb types are presented CLEC; however, this is not true if we consider the sizes of the two corpora, which will be shown by their type-token ratios. The type-token ratios are 8.9% and 24.9% for the Chinese learner corpus and the native

corpus. Again, the native students are demonstrated to use wider varieties of PVs with OUT.

**Table 7.1: Top five most frequent verbs in CLEC and LOCNESS**

<b>Rank</b>	<b>CLEC</b>	<b>LOCNESS</b>
1	GO (301)	<b>CARRY (65)</b>
2	<b>FIND (120)</b>	POINT (41)
3	<b>CARRY (100)</b>	<b>FIND (25)</b>
4	PUT (60)	GO (17)
5	JUMP (57)/ TAKE (57)	GET (14)

Table 7.1 above displays the top five most frequent verbs in their rank order, with the absolute frequencies in brackets. Three verb types: CARRY, FIND and GO, are the same in the two corpora. As found in the previous chapter, the over-/under-representations of PVs are not necessarily useful for a detailed analysis; as such, this will not be discussed further in this chapter. In the next section, we will concentrate on individual verb types for qualitative analysis. CARRY OUT and FIND OUT are selected for further examination because of their high frequencies and prevalence in the two corpora (Both of these PVs rank in the top three, as shown in Table 7.1). The other frequent PVs, GO OUT (top in CLEC) and POINT OUT (second in LOCNESS) are discarded, since each is only prominent in one corpus.

### **7.3 The case of CARRY OUT**

Many studies of learner corpora are designed to capture the features that differentiate between learners and English native users. Among those, the targets dealt with by the studies fall into three broad categories: (1) errors; (2) phraseology, chunks, prefabricated patterns, lexical and grammatical patterns; (3) colligation, or tag sequence. In this section, the aim is to answer the question as to how NS and NNS students of English differ in using one particular phrasal verb at the lexico-grammatical level, either within the word boundary or beyond. I will study the relations between individual words and more abstract units such as meanings or concepts. For the former (word relations), the verb will be examined by its word forms and collocation; for the latter (concept relations), an extensive analysis is employed to bring to light the semantic sequences involved in the usage of the verb.

I have attempted to explore the possible ways to best describe learner language. The previous analysis of PVs with UP suggests that a numerical study can point out interesting examples of PVs but reveals little about the usage. In addition, this also shows that each PV has unique behaviours, thus cannot be accounted for by taking all of the PVs as a whole. A better solution, then, may be to narrow down the study to particular cases of PVs. This will be tested by focusing on one PV, say CARRY OUT.

We will begin with the analyses of different word forms of CARRY OUT found in the NS and NNS texts. The collocates of CARRY OUT are also identified and grouped by their semantic fields. The distribution of these semantic groups indicates the users' perception of what items are of particular relevance to CARRY OUT. One collocate, *law(s)*, is studied and leads us to an approximate equivalent of the PV: ENFORCE. This verb is compared with CARRY OUT to mark out the differences. A number of uses of CARRY OUT exclusively utilised by the NSs are also brought into view. This section finishes with a thorough analysis of the co-occurring items, which include fixed elements like words or phrases and patterns beyond words. The combinations of these elements will be represented by semantic sequences.

### **7.3.1 Lexico-grammatical analysis**

#### ***7.3.1.1 Word forms***

The frequencies of each word form of CARRY OUT are listed in Table 7.2. The relative frequencies are the frequencies per million words, and the percentages are the proportion of the word forms in all cases of CARRY OUT in each corpus. An overview of the figures reveals an unexpected similarity in the proportions between each word form in both corpora. This may be because the Chinese learner and native speaker usage is the same, or it may be that a similarity in numbers is masking

difference. Further analysis can reveal the difference. For example, the highest percentages of the -ed word form suggest that NNSs and NSs favour this type most, and both NSs and NNSs utilise *carried out* to the same extent, around 50%. However, in fact, 88% of the NS data are used in passive voice, but in the NNS data, many examples of *carried out* are the past tense or perfect aspect, and only 75% are in the real passive. As such, a caveat can be made that reliance on numbers of surface forms may be less justifiable if further analysis (e.g. grammatical or semantic functions) is not performed at the same time.

**Table 7.2: Frequencies of word forms of CARRY OUT**

	CLEC			LOCNESS		
	abs.	rel.	%	abs.	rel.	%
<i>carry</i>	32	32	32	19	58.6	29.2
<i>carried</i>	50	50	49	32	98.7	49.2
<i>carries</i>	3	3	3	5	15.4	7.7
<i>carrying</i>	15	15	15	9	27.8	13.6
TOTAL	100		100	65		100

### 7.3.1.2 Collocation

Combining the appropriate collocation of a verb is also a crucial part of language knowledge for learners. Now I will examine the collocates of CARRY OUT which

play certain semantic roles (agents, patients and themes, as mentioned in Chapter 5), and summarise the results in Table 7.3. There are instances whose agents or patients/themes cannot be traced (e.g. *Carrying it out genetically would give more control*. [LOC]), and these will not be considered.

As regards agents, the NNSs and NSs are found to share similarity. Both of them tend to use nouns which construe a specific collection (e.g. *universities, nations, government*, etc.). It appears that the Chinese learners and native writers usually assume the agent to be an authority which has the right, power or group forces to execute something. If we move on to the words as patients/themes, dissimilarity is observed. The large number of the 4<sup>th</sup> category ‘instructions or requests’ in CLEC indicates that the words as patients/themes usually involve some public affairs (*laws, policy...*), which are to be put into effect. In contrast, the native speakers use more words of the 1<sup>st</sup> category, ‘actions’, and 2<sup>nd</sup> category, ‘activities’, signifying that CARRY OUT might be more associated with the implementation of an action/activity rather than the execution of public affairs.

**Table 7.3: Occurrences of agents and patients/themes across corpora**

LOCNESS					
Agent			Patients/themes		
Category	Example	%	Category	Example	%
1.(special name of) a person or an organisation	<i>Voltaire, PCF</i>	36.4	1.actions	<i>murder, assassination, massacre des Innocents deed, violence, reign of terror</i>	35.9
2.pronouns	<i>he, we</i>	27.3	2. activities	<i>task, treatment, studies, research</i>	26.6
3.a group of people of shared characteristics	<i>doctors, criminals, government, the higher class</i>	30.3	3.abstract notions	<i>function, the impossible</i>	4.7
4.others	<i>brain, anyone</i>	6	4.instructions or requests from others	<i>orders, wishes, policies</i>	15.6
			5. pronouns	<i>it</i>	10.9
			6.others		6.3

CLEC					
Agent			Recipient		
Category	Example	%	Category	Example	%
1.(special name of) a person or an organisation	<i>China, Kuwaiti</i>	9.8	1.actions	<i>crime, measures, euthanasia</i>	18.8
2.pronouns	<i>we, he</i>	45.1	2. activities	<i>procedure, assessment, reform</i>	21.9
3.a group of people of shared characteristics	<i>universities, countries, government, nations, societies</i>	45.1	3.abstract notions	<i>self-value</i>	2.1
4.others	--	0	4.instructions or requests from others	<i>laws, policy, principles</i>	38.5
			5. pronouns	<i>it</i>	5.2
			6.others	<i>everything</i>	13.4

Note: The calculation of the percentages is carried out thus: the numbers of each category are divided by the total number of 'CARRY OUT' in each corpus, excluding those without agents or patients/themes.

To explore the dissimilarity between the two corpora further, one of the collocates, *law(s)*, was studied further. The noun *law(s)* is selected because it frequently occurs in CLEC, while it is absent in LOCNESS. Since this might result from the small size of LOCNESS, BoE was consulted instead. *Law(s)* is apparently not one of the most frequent nouns that collocate with CARRY OUT returned in BoE. That is to say, *law(s)* and CARRY OUT are not strongly collocated in general English. The examples of *law(s)* + CARRY OUT generated from BoE are listed as [7-1] to [7-3] below, and at the points where similar meaning is expressed, the verb ENFORCE is preferred and occurs much more than CARRY OUT (citations [7-4] to [7-6]). By contrast, in sentences where laws are suggested to be put into effect, the Chinese learners seem to fail to convey the ideas they intended by using the precise verb (citations [7-7] to [7-13]), although in some cases it may be arguable that the use of CARRY OUT also makes sense (e.g. citations [7-12], [7-13]). The tendency of combining CARRY OUT and *law(s)* by the Chinese learners can be a result of L1 influence, because CARRY OUT + *law(s)* can be glossed in Chinese as an idiomatic two-word verb, ‘zhǐ fǎ’.

It may appear that the Chinese learners do not use ENFORCE with *law(s)*. In

fact, a follow-up search shows that some Chinese learners have no problem using this verb, but the citations are all from ST5-ST6, subcorpora of advanced college students (English-major) (citations [7-14] to [7-16]). To summarise, the Chinese learners, especially those who are non-advanced, are prone to confusing the two words, because they do not know that CARRY OUT collocates more often with actions such as *work, attacks, threats*, but ENFORCE tends to co-occur with social conditions like *laws, rules, ban*, etc. (examples are cited from BoE). These two verbs are often regarded to be synonymous to some extent; for example, ENFORCE is glossed as ‘to carry out effectively as in *enforce laws*’ in the Merriam-Webster Dictionary (2003). However, they have a subtle sense difference as ENFORCE emphasises compelling the observance of laws or rules while CARRY OUT suggests only execution. These two verbs draw our attention to distinguishing the usage difference between near-synonyms, which will be studied further with the examples FIND and FIND OUT later.

[7-1] The term executive branch suggests the branch of the federal government that executes or carries out the law. [BoE]

[7-2] In the UK the judiciary carries out the law as laid down by the legislature (Parliament). [BoE]

[7-3] the referee has to carry out the law and you can't blame him. [BoE]

[7-4] Governor Hutchinson...was equally determined to collect the tax and **enforce** the law.  
[BoE]

[7-5] Despite such abolitionist resistance, the Fugitive Slave Law was **enforced** fairly successfully in the early 1850s. [BoE]

[7-6] President Washington determined 'to go to every length that the Constitution and laws would permit' to **enforce** the law. [BoE]

[7-7] If the government still take [vp3,2-] no action to carry out the law to legalize it, euthanasia will be taken advantage by the criminals. [CLEC]

[7-8] Laws of environments were passed and carried out in many countries. [CLEC]

[7-9] taking [vp7,s-] to stop the deterioration of environment. New laws are made and carried out.[CLEC]

[7-10] All these event [np3, 1-] could be avoided if a clarified law had been carried out in China.[CLEC]

[7-11] From my foint [fm1,-] of view, the fake commodities must be got rid of. At present, many people have devoted themself [fm2,-] into [wd3,2-1]the action. Futhermore [fm1,-], I think that a law must be carried out to prevent [cc3,-2] the phenomena. 2,-] into [wd3,2-1]the action.[CLEC]

[7-12] Singapore has won the world reputation of the most perfect state in carrying out the laws for at least a decade. [CLEC]

[7-13] Fake [fm3,-] commodities, because our laws are not carried out efficiently.[CLEC]

[7-14] Its slackness in **enforcing** the laws is a fatal weakness. [CLEC]

[7-15] We should not **enforce** one law on our own citizens...[CLEC]

[7-16] To do that, the government has to **enforce** a high tax law. [CLEC]

Some citations used by the native students merit further discussion. The British and American students create sentences like those shown below (citations [7-17] to [7-23]). The underlined collocates are all abstract nouns. From Table 7.3 we can see that the Chinese learners are capable of dealing with some abstract nouns, but the possibility that the Chinese learners will match those nouns with CARRY OUT seems to be low, because these combinations do not often occur in the Chinese system. None of these usages were found in CLEC, suggesting that the Chinese learners may not be familiar with the combinations of CARRY OUT and words like *role*, *pregnancy*, *reign*, *justice*, *violence*, etc., which do not attract mutually in their L1. If these combinations are to be translated into Chinese, the phrasal verb CARRY OUT will be substituted by various Chinese verbs. As such, the combinations which are incongruent in English and Chinese are likely to be treacherous for the Chinese learners, and are evidences of

the Chinese learners' lack of full lexical knowledge which result in a limited collocation range.

[7-17] he said that he intended to carry out the role of president to the full extent of his powers

[7-18] Why can't these women just carry out the pregnancy and put the child up for adoption.

[7-19] Is Caligula right or wrong in carrying out his reign of terror?

[7-20] As if some justice is being carried out in the equal distribution of wealth,

[7-21] However Kaliyev carries out his political violence for the good of other people.

[7-22] One common denominator however is that the violence is carried out in the name of the communists.

[7-23] and the women are prepared for the job, that political violence can be carried out by anyone.

In addition, it is interesting to discover if the dictionaries capture these abstract nouns. Therefore, I checked two PV dictionaries, CPVD (2006) and MPVP (2005) and one learner dictionary, CALED (2006), finding that the connections linking the PV and these abstract nouns are totally omitted. In other words, the above combinations may not be revealed to learners. The dictionaries could benefit the

foreign students more if they provided more refined explanations of the senses and usages of a phrasal verb.

### 7.3.1.3 *Semantic sequences*

I took a further step to extend the investigation to longer phraseological patterns, such as semantic sequences. That is, I will look at the various kinds of ‘functions’ or ‘concepts’ (or ‘semantic labels’/‘discourse functions’ in Hunston’s (2008) terms) which occur with CARRY OUT in a bundle. As demonstrated by the examples of *the observation + that-clause + [consistency] + [theory/argument]* (Hunston, 2008:279), *[logical basis] + it is clear that +[claim]* (Charles, 2004) and so on, the patterns identified can comprise words, phrases or clauses, which constitute a unit in order to realise certain functions or express certain meanings. More importantly, they come together to help to shape the meaning of the verb. These semantic labels/discourse functions/concepts (e.g. [logical basis]) cannot be obtained directly from the exact words but can be deduced by the abstraction of these words. Also, the order of their elements can be indeterminate or flexible.

I searched for any such patterns in the corpora but no recurrent expressions were captured in CLEC. However by contrast, one special phraseological pattern, CARRY OUT + *on* was found to recur in LOCNESS. Six out of the 65 occurrences of CARRY

OUT contain this pattern. Apart from citation [7-29], which is an idiom *on impulse*,

all of the other five fall into a semantic sequence as below:

[7-24] much research is being carried out on human genes to find out how they work

[7-25] but nowadays biological operations can be carried out on humans which are not even born.

[7-26] Capital punishment has been carried out on criminals for quite some time,

[7-27] All the above are carried out on living people,

[7-28] Certain studies are carried out on people of one race.

[7-29] Punishments for violent actions that are thought out or planned rather than carried out on impulse.

<b>events</b>	<b>be carried out on</b>	<b>entities undergo the changes</b>
usually tests, research ...	usually passive voice	usually humans

The semantic sequence [events] + *be carried out on* + [entities undergo the changes] is often drawn on in the passive voice in order to lay emphasis on the entity which is greatly influenced by the outcome of particular events which have been or

will be carried out. Interestingly, it is noted that although some instances are neutral (e.g. citations [7-24], [7-28]), some of the uses of this phraseological pattern may suggest degrees of disagreement or censure, demonstrating the stance of the speaker (e.g. sympathetic). The nouns which follow the preposition *on* imply their roles as victims or disadvantaged groups, who are unable to defy (e.g. citations [7-25], [7-26], [7-27]).

This sequence is found to be absent in CLEC, providing further evidence indicating that the Chinese learners are less efficient at employing the full usage of this PV at their disposal. The consultation of BoE shows that the preposition *on* is the fifth most frequent word placed immediately after CARRY OUT. Evidently, this is an essential usage for the phrasal verb, but is largely overlooked by the Chinese learners, or the topics do not allow them to show this knowledge.

If cases which embed the phrase *carried out + on* are excluded, the remainder of the concordance lines of *carried out* in the two corpora also display interesting patterns. Certain types of semantic concept/function can be identified (some are ignored owing to their small numbers of occurrence or their being peripheral/non-core to the meaning construction of CARRY OUT; for example, concepts like **Time** and **Place** are left out in this study as theoretically they can be associated with many

verbs). Four conceptual elements are identified and respectively tagged as \_S (**Situation**), \_C (**Condition**), \_P (**Purpose**) and \_R (**Results**).

- **Situation:** The situation where an event happens or an action needs to be done. (e.g. *However, things have changed a lot when China began to carry out marketing economic system . [CLEC]*)
- **Condition:** The condition or constraint regarding how the event happens or action must be done. (e.g. *The Chinese people will try their best to carry out their lighter [wd3,-] future led by the Chinese Communist Party. [CLEC]*)
- **Purpose:** The goal or purpose why the action should be taken. (e.g. *So harmful are the fake commodities, thus we must carry out ways to deal with them. [CLEC]*)
- **Result:** The consequence or outcome which will be caused by the event or action. This can be either positive/beneficial or negative/undesirable. (e.g. *Since many countries carry out the economic reforms, the people's living standards are higher and higher. [CLEC]*)

**Table 7.4: Proportional distribution of ‘concepts’ with ‘carried out’**

	CLEC		LOCNESS	
	occurrence	%	occurrence	%
Purpose	4	8	5	15.6
Result	18	36	3	9.4
Condition	5	10	4	12.5
Situation	3	6	2	6.3

In Table 7.4, the percentages are based on the total numbers of *carried out* lines in the two corpora (N=50 in CLEC and N=32 in LOCNESS). The greatest difference is shown in their uses of the **Result** element. The occurrences of a **Result** element in CLEC overwhelm those found in LOCNESS. As a matter of fact, the Chinese learners incline towards a usage pattern which sets up a cause-and-result relationship. The emphasis is put on the consequence which results from carrying out the action intended. Some typical examples include:

[7-30] We have got great achievements since we carried out this policy in 1978.

[7-31] Since the universities carried out these reforms, they have provided more and more students who can work practically and effeciently [fm1,-] in the real work.

[7-32] Then after a crime is carried out, there are two possible result [np3,2-] before you.

[7-33] Since China carried out Reform [fm3, 1-1] and Open-Door [fm3, -1]Policy, great changes have taken place.

[7-34] In such a social environment [fm1,-] where laws and rules were properly obeyed and carried out, people's living standard improved, economy grew, political life was stable, which resulted in a real civilized nation.

With only 9.4% including the result element, clearly this phenomenon is not commonly considered in native writings. It appears to be a special feature of the Chinese learner language, in which this pattern is expressed quite consistently by the Chinese learners. Again, the NNSs are found to exhibit a very different way of presenting their thoughts from the NSs.

This finding and the earlier observation regarding collocates, that the NNSs often relate the execution of public affairs by authorities to this PV, together bear out the idea that the Chinese learners may be constrained in a fixed mode or stereotype. This may be caused by cross-language difference (the Chinese-English difference) or the effect of instruction.

To recap, it is ascertained that the Chinese learner language displays different behaviours as compared with native English. So far we have seen that the Chinese learner language differs from the native standard in a number of aspects such as

different word form distribution, concentration on collocates of other semantic groups, and an absence of important formulaic sequences or selections of biased semantic sequences. It is also noted that the Chinese learners have problems with a PV and a virtually equivalent single-word verb. In the upcoming section, I will turn to another example, FIND OUT, and compare it with its near-synonym, FIND.

## **7.4 The case of FIND OUT**

### **7.4.1 Introduction**

In this section, I will probe the different patterns of the phrasal verb FIND OUT used by the NNSs and NSs. In addition, the patterns of FIND will also be scrutinised, as this single-word verb overlaps with the phrasal verb FIND OUT to a great extent, bringing extensive problems to the Chinese learners. As such, Section 7.4.2 will reveal the typical phraseological patterns of FIND OUT in native students' language and the different phraseological patterns in the Chinese learner language as compared with NS language. It begins with an analysis of the phraseological patterns following the PV. The predominant patterns will be revealed to demonstrate the disparity among NSs and NNSs. Subsequently, I will take a further step to look into the nouns which are frequently used with FIND OUT by the Chinese learners and divide them into

semantic types. This uncovers the lexical items which are most likely to be primed with FIND OUT for the Chinese learners. I then turn to the phraseological patterns preceding FIND OUT: the most dominant pattern and all the lexical items associated with this pattern are shown and compared with BoE. Furthermore, the differentiation of FIND and FIND OUT is tackled in Section 7.4.3. The phraseological patterns which are specific to FIND OUT and FIND in general English (i.e. BoE) will also be reported, followed by an examination of whether the Chinese learners successfully distinguish the synonymous FIND OUT and FIND.

## **7.4.2 Usage patterns of FIND OUT in CLEC, LOCNESS and BoE**

### ***7.4.2.1 Patterns following FIND OUT***

Five types of syntactic pattern that follow FIND OUT were identified and some examples are given: (1) Wh-words (e.g. *why, which, what*) and *how*: (*We would be able to find out where all of the money is going. [LOC]*) (2) *that*-clause (including the instances where *that* is omitted): (*When I got older I found out that my mom and my dad smoked weed when they were younger. [LOC]*) (3) VN (including pronouns): (*They try to find out their quick method. [CLEC]*) (4) *if*-clause: (*..., the most important thing for her was to find out if he had got married or not. [CLEC]*) (5) Miscellaneous

(e.g. *be V-ed*): (*They are nervous, they are afraid of being found out. [CLEC]*). The numbers and percentages of the five patterns across the corpora are presented in Table 7.5.

**Table 7.5: Distribution of following patterns of 'FIND OUT' in CLEC and LOCNESS**

	CLEC			LOCNESS		
	Raw Freq.	Nor. Freq.	%	Raw Freq.	Nor. Freq.	%
Wh-words and <i>how</i>	21	20	17.5	11	34	<b>44</b>
<i>that</i> -clause	20	19	16.7	2	6	8
VN	68	64	<b>56.7</b>	7	22	28
<i>if</i> -clause	2	2	1.7	3	9	12
Misc.	9	8	7.5	2	6	8
TOTAL	120	113	100	25	77	100

This table presents both the raw and normalised frequencies (per million words) and the percentages. The percentages are calculated based on the raw frequency rather than the normalised frequency because this is more accurate. We can either compare the normalised frequencies or the proportions of each pattern. The former tells us which pattern occurs more frequently in which corpus if the total numbers of word tokens are equivalent. The latter measures the proportions of the patterns in all cases of FIND OUT. Therefore, the proportions represent the distribution of these patterns

on the occasions where FIND OUT is used. The difference between these two approaches is that the normalised frequency approach takes account of the total number of word tokens in a corpus, while the proportion approach can be seen as a kind of relative frequency which is related only to all expressions of FIND OUT. Although the two approaches both show the five patterns ranking in the same order, it will tell a different story when each pattern is considered in its ratio across the corpora. For example, the ratio of the *if*-clause (LOCNESS: CLEC) in terms of normalised frequency is 9:2 (4.5 times), while it becomes 12:1.7 (7 times) in terms of percentage. We shall, therefore, be careful when interpreting these figures.

It seems to me the proportion approach is more appropriate for my purpose here, since I am comparing the occurrences of a pattern in relation to a particular PV (i.e. their co-occurrences). It is meaningful only when a PV is used, thus the comparison is less relevant to the total word tokens of a corpus. CLEC shows that FIND OUT is liable to be followed predominantly by a noun (VN) (56.7%), whereas LOCNESS shows more inclination towards the uses of *wh*-words (44%). The Chinese learners are also aware of the usage of FIND OUT + *wh*-words, since there are 17.5% uses of this pattern. However, this sequence is not as frequent in CLEC as in LOCNESS; the NSs use it more than twice as often as the Chinese learners (44% vs. 17.5%). This

suggests that this pattern is a typical usage of FIND OUT in NSs' language repertoire but the connection between the PV and *wh*-words may not be as strong in the Chinese learners' lexicon/phrasicon (The term 'phrasicon' first appeared in Fillmore (1978)). Whether this feature can be confirmed in general English will be tested later. Some citations of FIND OUT + *wh*-words found in LOCNESS are listed below:

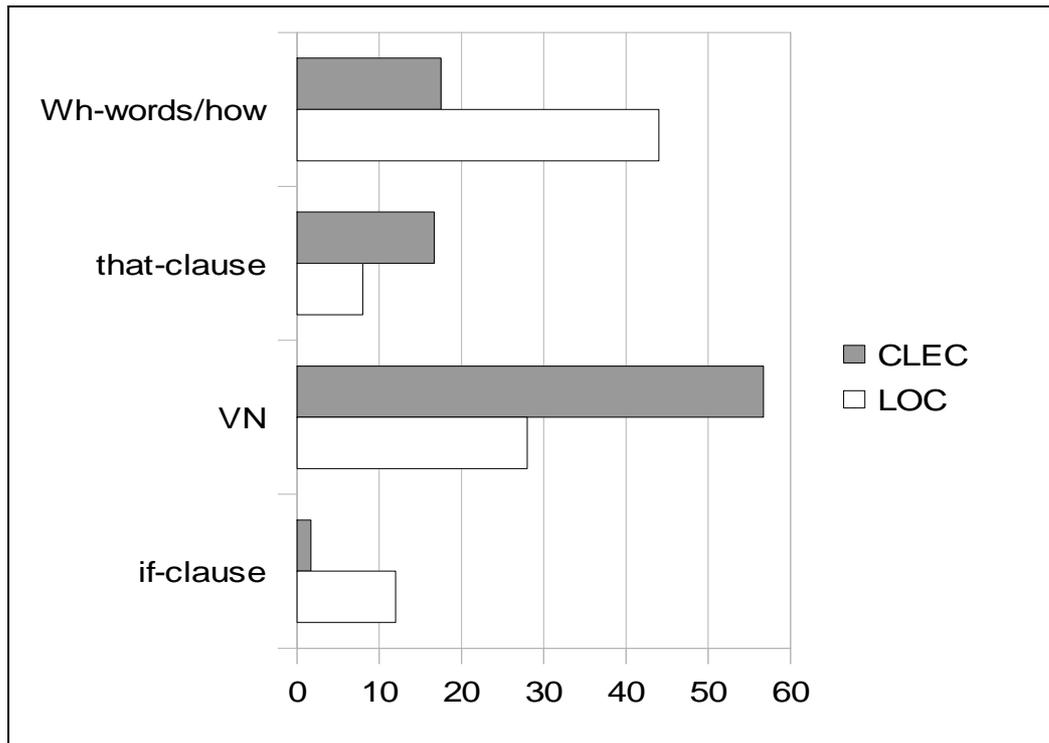
[7-35] The quest of **finding out who** one is is the quest that all individuals must embark...

[7-36] We would be able to **find out where** all of the money is going.

[7-37] ...they simply wanted to **find out what** it was all about...

In order to see the proportional contrasts of one pattern across corpora clearly, a bar chart is shown in Figure 5. This shows that the Chinese learners prefer the VN and *that*-clause patterns while the NSs prefer *Wh-words/how* and *if*-clause patterns. The differences between each pattern favoured in one corpus are generally around twice or more than in the other corpus (e.g. for the VN pattern, 56.7% is twice 28%), except for the *if*-clause pattern, which is seven times more frequent in LOCNESS. However, we need to be cautious here, since the raw frequencies of the *if*-clause are not sufficiently large to be convincing. In a sense, the *if*-clause is similar to *Wh-words/how*, because

they all indicate a question in an affirmative clause, and *if* and *whether* are interchangeable in most cases. Therefore, if we combine these two patterns into one, the ratios of percentages in CLEC and LOCNESS will become 19.2% vs. 56%, which again supports the idea that the *Wh-word/how/if* pattern is the most representative usage in respect to FIND OUT in L1 English, at least for British and American college students. This is also proved by investigating BoE, and the evidence will be shown later (the dominant patterns of FIND OUT in BoE are given in Section 7.4.3).



**Figure 7.1: Proportional distribution of patterns**

#### *7.4.2.2 The lexical associations of the VN structure*

The large amount of the VN pattern (56.7%) in CLEC gives us more confidence to make statements about the use of FIND OUT in L2 English. If we scrutinise the instances of VN in CLEC, we can gain a clearer picture of the subtle sub-senses of FIND OUT + N in the Chinese learner language. At first, the nouns following FIND OUT seem to be unclassifiable, since things which can be found out appear to be unrestricted, thus unlimited; however, to my surprise, the nouns form several clear and consistent categories, which are summarised below: the percentages are given in

parentheses and example collocates are listed in curled brackets. Note that the figures of the percentages are rounded, thus the total number is not precisely 100%. Those which do not fit into the first seven groups or are unidentifiable due to errors or context are put in the miscellaneous group.

- **something negative or harmful which needs to be located:** (20%) {mistake, murder, crime, drawback ...}, e.g. *we customers also should train the ability to find out fake commodities...[C]*
- **a solution:** (21%) {solution, answer, method ...}, e.g. *..., you should think and think to find out the best way.[C]*
- **a fact/ truth:** (10%) {truth, fact ...}, e.g. *But Eliza found the fact out herself.[C]*
- **something which has been lost:** (3%) {ticket ...}, e.g. *Grandpa Li found out Wanghua near wanghua's home.[C]*
- **new information/ discovery/ resource:** (13%) {gravitation, source, water, information ...}, e.g. *we can easy [wd2, 1-2] find out two trends about the healthy condition...[C]*
- **reason:** (12%) {cause, reason ...}, e.g. *...we should find out the reason we are short of fresh water.[C]*
- **something which is hidden:** (7%), e.g. *They began to find out the secret. [C]*

- **miscellaneous:** (16%)

Only seven instances of the VN group are found in LOCNESS. Although the number is small and may not represent the whole situation of the NS language, it might still be worth noting that there are two main patterns, FIND OUT + *truth* and FIND OUT + *about* + N in the VN group. Note that FIND OUT + *about* + N is usually not present as a continuous sequence; adverbs such as *more*, *a lot* often occur between the PV and *about*. Interestingly, the FIND OUT + *about* + N pattern is not present in the Chinese learner data. This raises the question whether FIND OUT + *about* + N is also a typical pattern in its usage. To answer this, BoE was consulted to discover the most frequent collocates of FIND OUT. The collocates obtained from BoE are presented in Table 7.6 in the order of their t-score. Words like *about*, *what*, *when*, *whether*, *why* and *how* are found frequently to co-occur with FIND OUT (these are emboldened). I examined the occurrences of these collocates and calculated their proportions of all FIND OUT entries. The results for each emboldened collocate are listed in Table 7.7. Obviously FIND OUT + *wh-words/how* and FIND OUT + *about* are characteristic patterns in the usage of FIND OUT. As we have seen earlier, the Chinese learners are found to be less likely to employ these two patterns: the former

pattern is not frequently used and the latter pattern is completely absent in CLEC.

The proportion 21.5% has again supported the idea that FIND OUT + *wh-words/how* is a very typical usage. The other pattern, FIND OUT + *about*, occurs in BoE about 10.5% of the time, which suggests that it is also an important pattern in respect to the usage of FIND OUT. Some citations of FIND OUT + *about* from BoE are exemplified below. The fact that this usage is missing in the Chinese learner data reminds us of the importance of covering all the possible patterns of a usage. Introducing the essential usages to the Chinese learners can help to enhance the variability and expressivity of their language.

[7-38] There are many ways to **find out about** ourselves and the world...

[7-39] Learning affects how we **find out about** the world and ourselves as...

[7-40] They have few ideas to **find out about** college majors.

[7-41] You can **find out more about** the scholarship...

[7-42] one ought to look to **find out about** the transformation of ...

**Table 7.6: The top 15 most frequent collocates of FIND OUT in BoE**

Rank	Lexical item	T-score
<b>1</b>	to	175.5
<b>2</b>	<b>what</b>	72.8
<b>3</b>	<b>about</b>	65.5
<b>4</b>	i	50.4
<b>5</b>	<b>how</b>	48.1
<b>6</b>	you	47
<b>7</b>	if	44.2
<b>8</b>	we	38.3
<b>9</b>	they	37.7
<b>10</b>	<b>when</b>	34.7
<b>11</b>	more	32.1
<b>12</b>	<b>whether</b>	28.9
<b>13</b>	<b>why</b>	28.8
<b>14</b>	she	27.9
<b>15</b>	way	27.2

**Table 7.7: Frequencies and percentages of some patterns in BoE**

<b>Pattern</b>	<b>Frequency</b>	<b>%</b>
FIND OUT + ( <i>more...</i> ) <i>about</i>	3414	<b>10.5</b>
FIND OUT+ <i>what</i>	3566	11
FIND OUT+ <i>when</i>	323	1
FIND OUT+ <i>whether</i>	796	2.5
FIND OUT+ <i>why</i>	674	2.1
FIND OUT+ <i>how</i>	1589	4.9
<b>SUBTOTAL</b>		<b>21.5</b>

It seems that the Chinese learners tend to use the first and second types of nouns (i.e. the ‘mistake’ group and the ‘solution’ group) with FIND OUT in the VN structure. Because they are not found in LOCNESS, we need to resort to BoE. However it is unfeasible, at this point, to investigate all of the concordance lines of FIND OUT in BoE and identify all of the nouns of these two types; an alternative approach must be adopted. In order to test whether this is a specific feature of Chinese learner language, the collocates *murder, crime, mistake* and *solution, answer, method* (and their plurals) from the top two groups (the ‘mistake’ and ‘solution’ groups) are selected for examination in BoE. The method is to query the strings which contain the PV and the different forms of the collocate so that the numbers of occurrences can be

obtained. The purpose is to examine whether the proportions of co-selecting these nouns with FIND OUT in BoE are larger than those found in CLEC.

The results are shown in Table 7.8. The total numbers of FIND OUT in CLEC and BoE are 120 and 32471 respectively. As all of the six collocates only occur once in CLEC, the percentages are all 0.8%, although the single occurrence may cast doubt on the reliability of whether the Chinese learners prefer to use them. This is justifiable when each collocate is taken as representing the group it belongs to (i.e. *murder, crime, mistake* represent the first group of negative things; *solution, answer, method* represent the second group of solutions). Compared with the percentages in BoE, the Chinese learners indeed seem to have a preference for employing the first and second noun types. It is arguable that the small percentages found in BoE may not suffice to prove this, because the size of BoE will naturally result in small percentages, but it is believed that this effect can be minimised, as the proportion is used here instead of the word tokens. The fact that the Chinese learners favour the use of negative things and solutions with FIND OUT is ascertained.

**Table 7.8: Frequencies and percentages of the six collocates with FIND OUT in BoE and CLEC**

	BoE		CLEC	
	Freq.	%	Freq.	%
<i>murder</i>	0	0.000	1	0.8
<i>crime</i>	1	0.003	1	0.8
<i>mistake</i>	5	0.015	1	0.8
SUBTOTAL	6	<b>0.018</b>	3	<b>2.5</b>
<i>solution</i>	5	0.015	1	0.8
<i>answer</i>	46	0.140	1	0.8
<i>method</i>	1	0.003	1	0.8
SUBTOTAL	52	<b>0.160</b>	3	<b>2.5</b>

Note: The percentage is calculated by dividing the frequency of the collocate with the total number of 'FIND OUT'.

#### 7.4.2.3 *The precedent patterns of FIND OUT*

I examined the context on the right hand side of FIND OUT as discussed earlier, and noticed that what precedes FIND OUT also displays certain patterns. The pattern 'to FIND OUT' is found most frequently on the left hand side of FIND OUT, both in the two corpora. I found that 36% and 37.5% of the instances in LOCNESS and CLEC contain this pattern respectively. No obvious extended patterns are noted in LOCNESS, but a semantic sequence 'it is + [evaluation] + to + FIND OUT' is found in CLEC.

Since the native students' data does not have this sequence, BoE was searched for comparison. I focused on adjectives which perform the [evaluation] function, since words of other classes rarely denote this function. All of the words found in the [evaluation] slot are listed in Table 7.9. The percentages represent the frequencies of each word out of the whole frequency of the sequence 'it is + [evaluation] + to + FIND OUT'. The [evaluation] words can be further classified into groups according to their similarity (semantic fields) as below. Each group is represented by the most typical word, and the function and members of each group are also listed (words in different groups are marked differently in Table 7.9).

- **The 'important' group**

- <function> to indicate the importance
- <members> *important, necessary, valuable, essential*

- **The 'easy' group**

- <function> to highlight the degrees of difficulties or easiness
- <members> *easy, difficult, hard*

- **The 'possible' group**

- <function> to express the possibility
- <members> *possible, impossible, likely*

- **The ‘best/better’ group**
  - <function> to suggest appropriate ways or evaluate the situations
  - <members> *best/better, useful, wise, advisable, helpful, worse*
  
- **The ‘surprising’ group**
  - <function> to show the feelings caused by the consequences or events  
  
found out
  - <members> *surprising, incredible, interesting, unnerving*

The Chinese learners clearly tend to use a fixed expression, *it is possible to FIND OUT*, although the number (11) is not large. These 11 sentences, however, come from different writers if the source texts are checked. *It is possible to FIND OUT* appears to be the most dominant or typical sequence available to the Chinese learners, suggesting that the elements in this sequence are primed (in Hoey's term) immediately when FIND OUT is used. After examining the data from BoE, we know that there are other semantic sequences which the Chinese learners can make use of, for example, *it is important to find out*, etc. It might be of pedagogical value for teachers to reveal to the students that there are other methods of expression in the classroom, or exemplify them in textbooks.

**Table 7.9: List of words in the EVALUATION slot**

CLEC			BoE		
Words	Freq	%	Words	Freq	%
possible	11	84.6	<b>important</b>	24	29.2
glad	1	7.6	<i>(not) easy/easier</i>	8	9.8
<i>difficult</i>	1	7.6	possible	7	8.5
			<i>difficult</i>	6	7.3
			<i>hard</i>	5	6.1
			impossible	5	6.1
			<b>necessary</b>	4	4.9
			best/better	3	3.7
			interesting	3	3.7
			<b>valuable</b>	3	3.7
			<i>(not/hardly)</i>	2	2.4
			surprising		
			<b>essential</b>	2	2.4
			useful	2	2.4
			wise	2	2.4
			advisable	1	1.2
			helpful	1	1.2
			incredible	1	1.2
			likely	1	1.2
			unnerving	1	1.2
			worse	1	1.2
			<b>TOTAL</b>	<b>82</b>	<b>100</b>

### 7.4.3 Comparing FIND and FIND OUT in CLEC and BoE

#### 7.4.3.1 Introduction

As shown, I have analysed the patterns and collocation of the phrasal verb FIND OUT in CLEC and LOCNESS/BoE. The Chinese learners are found to favour some usages which are not the most frequent ones in the native students' language. In the meantime, some particular uses appear to cause difficulties to the Chinese learners. Besides these discrepancies, one of the problems which particularly relate to FIND OUT is the overlap of this PV and its single verb counterpart. It is quite possible that the Chinese learners may confuse the usage of FIND OUT with FIND, since the verb and the phrasal verb are generally accepted as near-synonymous in English. The fine difference between FIND and FIND OUT is interesting because they are similar in both senses and word forms. They overlap in parts of the senses and the PV is an extension of the single-word verb in the form.

Several studies have paid attention to the differentiation of near-synonymous words. For example, Kennedy (1991) investigates *between* and *through*, and Biber, Conrad and Reppen (1994) look into *certain* and *sure*. These studies all successfully identify the grammatical or semantical usage differences of the near-synonyms, highlighting the importance to language learners of distinguishing near-synonyms.

This section (7.4.3) thus aims to differentiate the usages of FIND and FIND OUT in general English, and investigate the Chinese learners' uses of these two verbs. The background of their sense differences/similarities will be provided first by resorting to dictionaries and the literature, followed by a report on the analysis of the individual specific patterns of FIND OUT and FIND. The usage patterns of FIND OUT are laid out in detail here to illustrate the points mentioned in Section 7.4.2, where certain pieces of evidence from general English are required. Evidence on FIND from BoE will also be listed thoroughly in comparison with the PV. Moreover, form distribution and different approaches to examining usage differences of these two verbs will be presented. Next I display the results of investigating the Chinese learners' misuses and problems, then end with a brief summary of the findings regarding FIND OUT and FIND.

#### ***7.4.3.2 Confusion of FIND and FIND OUT***

At this point, it is necessary to review how the two synonymous verbs are introduced to learners in the reference books. The differences/similarities of these two verbs can be identified by looking up their definitions in dictionaries, therefore FIND and FIND OUT were probed in three dictionaries which are commonly consulted by foreign language learners: *Collins Cobuild Advanced Learners' English Dictionary* (2006),

*Cambridge Advanced Learner's Dictionary* (2005), and *Merriam-Webster Collegiate Dictionary* (2003) (see Appendix D for the sense list). In these three dictionaries, FIND has many different senses while FIND OUT usually has fewer. Though the entries of different senses are not consistent in these three dictionaries, we can still discern the major senses of these two verbs. It appears that, according to the three dictionaries, FIND OUT has two main senses: 'to learn/realise or discover' and 'to catch or discover someone/something wrong/dishonest', whereas FIND has many sub-senses derived from the primary sense 'discover or realise/learn'. The primary sense 'discover or learn/realise' is shared both by FIND and FIND OUT, as seen in the dictionaries. Not only is the meaning similar, but the major usage patterns are also the same; both patterns VN and V *that* can be used with the sense 'discover or learn/realise'. This is the area where these two verbs are found to be synonymous to a great extent. This causes confusion for learners and leads them to regard these two verbs as being freely interchangeable. However, such an interchange is dangerous, as there are some usages which are specific to one verb but not the other. For example, this can lead to the errors, *\*fiber is found out in cereal* (in the sense of 'exist'), *\*the culprits were soon found* (in the sense of 'identified' or 'to catch in an offence' but not 'discover in somewhere'). Learners should be aware of what expressions are allowed

for one verb but not the other, and vice versa. A secondary reason for confusion comes from the need to establish the associations between the structure and meaning, as indicated by Sinclair (1991). The sub-senses of FIND are determined by the words surrounding it, for example, FIND is glossed as ‘to become aware of’ when it co-occurs with a description of situations as in *We came home to find the cat had had kittens* (Walter, 2005). Unfortunately such kinds of link are not explicitly presented, and the distinctions of the single-word verb and PV are also not made clear in these dictionaries, so learners may suffer confusion when acquiring knowledge from them. As such, I will adopt a corpus approach to identify the distinctive phraseological patterns as regards FIND and FIND OUT later.

Another aspect of confusion with FIND and FIND OUT is the addition of the particle. The particle is regarded to add new meaning to the verb in cognitive linguistics, where the particle OUT is deemed to suggest ‘becoming seen or know’ (Lam, 2003:121) and ‘existence, knowledge, visibility, availability’ (Neagu, 2007:129). These statements imply that the distinction between a verb and a phrasal verb (the same verb with OUT) is that the phrasal verb indicates revelation or discovery but the verb does not (e.g. *turn* and *turn out*). However, this assumption cannot account for the usage distinction of FIND and FIND OUT, since it fails to take

into account the fact that the verb FIND itself already carries a similar meaning of revelation or discovery, which overlaps with the meaning suggested by OUT.

#### ***7.4.3.3 Comparison of FIND and FIND OUT in BoE***

As seen already, the two synonymous verbs are not well accounted for in the dictionaries and literature; attention should be paid to the empirical data showing the behaviours of FIND and FIND OUT before we can move on to observe whether the Chinese learner language has similar performance. In the respect that different word forms of a lemma may have different behaviours, the investigation will start with the distribution of word forms.

The statistics for the word forms of FIND and FIND OUT in BoE are presented in Table 7.10 and Table 7.11. The examples of FIND OUT were weeded out from the numbers of FIND to make sure that the occurrences of FIND do not include the phrasal verb. The percentages are the proportions of the occurrences of each form divided by the total occurrences of FIND (total= 286095) and FIND OUT (total=28861). The numbers for the simple verb FIND are roughly 10 times more than the phrasal verb FIND OUT. Therefore we know that PVs are far less used than their single-verb counterpart in English by nature, as evidenced by the distinction in this example. Overall, the distribution of the word forms is similar in both groups. For

example, the base forms *find* and *find out* are the most dominant forms used.

**Table 7.10: Forms of FIND in BoE**

<b>Form</b>	<b>Freq</b>	<b>%</b>
<i>find</i>	152778	53.4
<i>finding</i>	23964	8.4
<i>finds</i>	15532	5.4
<i>found</i>	93821	32.8
<b>TOTAL</b>	<b>286095</b>	

**Table 7.11: Forms of FIND OUT in BoE**

<b>Form</b>	<b>Freq</b>	<b>%</b>
<i>find out</i>	19472	67.5
<i>finding out</i>	1987	6.9
<i>finds out</i>	891	3.1
<i>found out</i>	6511	22.6
<b>TOTAL</b>	<b>28861</b>	

The phraseological patterns of the two synonymous verbs are discovered at this point. To extract the phraseological patterns from BoE, the ‘Picture’ function is used and results are shown in the tables below, in the order of the lemma, the present tense form, and the past/perfect form. Collocates which are considered to form interesting patterns are emboldened.

**Table 7.12: Patterns of FIND OUT in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	you	<b>to</b>	<b>FIND OUT</b>	out	<b>what</b>	the
to	<b>trying</b>	and		a	<b>how</b>	<b>about</b>
and	we	can		<b>that</b>	more	they
i	<b>try</b>	you		<b>themselves</b>	<b>if</b>	you
you	<b>want</b>	ll		this	<b>about</b>	it
we	<p>	will		it	the	out
a	i	i		<b>yourself</b>	<b>who</b>	he
be	<b>way</b>	t		<b>himself</b>	<b>whether</b>	s
was	and	could		the	<b>why</b>	was
in	can	we		their	that	i

**Table 7.13: Patterns of the word form ‘find out’ in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	you	<b>to</b>	<b>find out</b>	out	<b>what</b>	the
to	<b>trying</b>	and		a	<b>how</b>	<b>about</b>
and	we	can		<b>that</b>	more	they
i	<b>try</b>	you		<b>themselves</b>	<b>if</b>	you
you	<b>want</b>	ll		this	<b>about</b>	it
we	<p>	will		it	the	out
a	i	i		<b>yourself</b>	<b>who</b>	he
be	<b>way</b>	t		<b>himself</b>	<b>whether</b>	s
was	and	could		the	<b>why</b>	was
in	can	we		their	that	i

**Table 7.14: Patterns of the word-form ‘found out’ in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	<b>when</b>	i	<b>found out</b>	out	that	the
and	i	he		<b>that</b>	<b>about</b>	was
but	and	we		<b>himself</b>	out	out
i	he	they		this	i	it
<b>when</b>	<b>if</b>	she		<b>themselves</b>	the	had
as	the	have		a	he	he
<p>	we	and		him	what	i
<b>if</b>	as	be		<b>myself</b>	she	of
that	that	<b>only</b>		it	and	that
s	have	've		<b>herself</b>	it	she

**Table 7.15: Patterns of the word-form ‘found out’ in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	you	<b>to</b>	<b>FIND</b>	<b>a</b>	<b>in</b>	to
and	i	i		<b>out</b>	the	the
you	we	will		the	<b>way</b>	<b>of</b>
i	they	you		it	what	in
to	<b>trying</b>	can		<b>that</b>	to	and
that	and	ll		<b>themselves</b>	a	a
is	can	t		their	of	for
was	he	and		them	and	that
a	will	could		<b>in</b>	more	you
of	<b>if</b>	we		an	<b>hard</b>	s

**Table 7.16: Patterns of the word-form ‘find’ in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	you	<b>to</b>	<b>find</b>	<b>a</b>	<b>in</b>	to
and	i	i		<b>out</b>	the	the
you	we	will		the	<b>way</b>	<b>of</b>
i	they	you		it	what	in
to	<b>trying</b>	can		<b>that</b>	to	and
that	and	ll		<b>themselves</b>	a	a
is	can	t		their	of	for
was	he	and		them	and	that
a	will	could		<b>in</b>	more	you
of	<b>if</b>	we		an	<b>hard</b>	s

**Table 7.17: Patterns of the word-form ‘found’ in BoE by frequency**

<b>L3</b>	<b>L2</b>	<b>L1</b>	<b>Node</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>
the	the	i	<b>found</b>	<b>that</b>	the	the
and	to	be		<b>in</b>	in	to
of	and	was		the	a	of
a	i	have		a	of	in
in	can	he		it	to	a
that	he	and		out	be	and
<p>	have	they		to	and	was
s	they	been		<b>himself</b>	that	s
but	has	has		<b>guilty</b>	<b>way</b>	that
to	when	were		<b>on</b>	on	on

The phrasal verb FIND OUT, with its two forms *find out* and *found out*, was treated first, yielding a number of significant patterns, as shown from Table 7.12 to Table 7.14. The first pattern we can note is that FIND OUT is followed by a number of *wh*-words, as already observed in LOCNESS in Section 7.4.2. Moreover, the phrasal verb is commonly positioned after certain volition verbs such as TRY, WANT (also MUST, NEED, *would like to*, etc. in lower ranks). FIND OUT also co-occurs frequently with *about* and *oneself* (including *themselves*, *yourself*, *himself*, etc.). A probe of the form *find out* returned exactly the same result as FIND OUT, thus will not be repeatedly discussed. Unlike *find out*, examining the form *found out* reveals three typical patterns: one is also presented in the FIND OUT data (*found out+oneself*), and two different patterns (*when/if*, *only*) solely belong to *found out* as prominent patterns.

I then move on to analyse the results of FIND from Table 7.15 to Table 7.17. Note that the results of FIND include FIND OUT as well, because unfortunately the programs in BoE cannot separate them, thus requiring more careful interpretation. By the same token, FIND is analysed with the forms *find* and *found*. Unsurprisingly, some similar patterns to those found with FIND OUT also arise with FIND. For

example, FIND is also closely linked with the volition verbs TRY (and WANT in lower rank). This is fairly possible partially due to the inclusion of FIND OUT in the raw data, as approximately 23% of the examples of *trying to find* are in fact *trying to find out*<sup>vi</sup>. Excluding these 647 lines of *trying to find out* leaves 2347 lines of *trying to find* (not followed by *out*), proving that *trying* remains an essential collocate of FIND.

The reflexives are also fairly frequent collocates of both FIND (*themselves, myself, himself, etc.*) and FIND OUT. In addition, for other predominant patterns of FIND such as FIND + *in* (rank 1st, R2 and rank 9th, R1) and FIND + *way* (rank 1st, R2), the preposition *in* (rank 15th, R2) and the noun *way* (rank 22th, R2) also occur together with FIND OUT. However, one collocate, *hard*, is found to be popular with FIND but not FIND OUT, along with other synonymous words such as *difficult* (rank 11th, R2 and rank 26th, R3) (The sequence is FIND something *hard/difficult*). The data for *find* is also the same as that arising for FIND, and will not be further discussed. When it comes to *found*, a strange pattern is noted to be prevalent, that *to* precedes *found*, that is *to found*. Scrutinising the concordance lines shows that, although they are abundant, these just happen to be native speakers' deviated performance and will not be discussed further. Besides the anomalies, the collocates of *found* do not appear to have distinctive patterns except one unique word *guilty*, which constitutes a special

expression *be found guilty*. These findings are given in Table 7.18 below for a clearer view.

**Table 7.18: Summary of typical patterns of FIND OUT and FIND**

Node words	Patterns
<b>FIND OUT</b>	FIND OUT+ <i>what/how/if/who/whether/why</i> <i>trying/want to</i> +FIND OUT FIND OUT+ <i>about</i> FIND OUT+ <i>oneself</i>
<i>find out</i>	Same as FIND OUT above
<i>found out</i>	<i>when/if+found out+that</i> <i>found out+oneself</i> <i>only+found out</i>
<b>FIND</b>	<i>trying to</i> +FIND FIND + <i>way of</i> FIND+ <i>in</i> FIND+ <i>hard</i>
<i>find</i>	Same as FIND above
<i>found</i>	<i>found+guilty</i>

These findings demonstrate that most patterns are shared by both the single-word verb and the phrasal verb, except a few patterns which can be said to characterise one verb, although they may not be exclusive to one verb only. In the following section, I will investigate certain patterns which tend to belong to only one of the two near-synonyms in the Chinese learner corpus to examine whether Chinese learners confuse their usages.

The frequencies of FIND OUT retrieved from CLEC and LOCNESS are 109

times and 77 times per million words respectively; however, FIND is retrieved 1941 times per million words (freq = 2078) in CLEC and 931 times per million words (freq = 302) in LOCNESS. Apparently the numbers of FIND overwhelm these of the phrasal verb both in CLEC and LOCNESS, as was also found in BoE.

Certain patterns shown in the previous section which are strongly associated with FIND but not FIND OUT will be examined at this point. Because from the senses listed in dictionaries, it is noted that generally the usages of FIND cover a great extent of FIND OUT, it may be more fruitful to research FIND rather than FIND OUT. Studying these specific usages helps us note whether the Chinese learners misuse them because of their confusion of these two verbs. It is not my intention to give an exhaustive list of all of the possible specific usages, but to select only a few examples to account for the differences between FIND and FIND OUT. We have seen their typical patterns from the BoE data earlier, from which two patterns specific to FIND are selected for further research as shown below:

- FIND+(N)+Adj+(to) /be found+Adj
- FIND+(no)+Adj+N

The first is FIND+(N)+Adj, with some sub-types, such as FIND+N+Adj+ *to*, *be found*+Adj. Examples are FIND *it difficult to* (5%), *be found guilty*. This is a usage

which never happens with FIND OUT, e.g.: *\*find it out difficult to*, *\*be found out guilty* (as evidenced by BoE). Another usage which only occurs with FIND but not the phrasal verb is FIND+*no*+(Adj)+N. Citations [7-43]-[7-46] below are examples of this pattern. The only instance found in BoE where FIND OUT is followed by *no* is actually a *that*-clause (citation [7-47]).

[7-43] We **find no** evidence, for example, that...

[7-44] ...where he would **find no** dinosaur skeletons.

[7-45] The survey **found no** major difference.

[7-46] ...the American Medical Association **found no** heightened risk of cancer...

[7-47] ...before you arrived soon wears off when you find out no-one else has read their either.

In all of the citations of FIND OUT in CLEC, none was found to be used in these two patterns, suggesting that the Chinese learners do not appear to confuse these two usages. However, we should not jump to the conclusion that the Chinese learners have no problems in using these two synonymous verbs at all, since only two patterns have been probed. Further research to include more patterns is required before any hasty

interpretation can be made.

Since the study of the specific patterns did not reveal much tangible evidences of the Chinese learners' confusion of the single-word verb and the phrasal verb, I turn to look at the concordance lines one by one in order to identify instances that reveal the Chinese learners' problems with ambiguity and confusion of FIND and FIND OUT. The problematic cases are those which may be perceived as 'errors' or 'mistakes' because the usage is rather unacceptable in English. They are shown in examples [7-48]-[7-56] below. The first two instances are used in the sense of 'come upon/ learn where someone/something is', thus FIND is more appropriate. Therefore, the Chinese learners appear to be less aware that [place] is restricted with FIND but not FIND OUT. Citations [7-50]-[7-51] describe experiences perceivable by human senses, thus FIND is more suitable for use here. Again, [human senses] is more connected with FIND. For examples [7-52]-[7-54], the 'errors/mistakes' were recognised by checking the collocation of the verb and its object in BoE. If none or only an extremely small number of the collocations were returned, the instance is rendered 'unacceptable'. The use of the phrasal verb may not be taken as wrong, but consultation of BoE found only a few or no instances of the PV with the collocate, and at the same time a larger number with FIND were obtained. FIND OUT + *trend(s)* returned only 4 cases, all of

which have *about*, *what* before the object, but FIND was collocated with *trend(s)* 46 times. 'FIND OUT + *drawback(s)*' does not exist in BoE, but FIND + *drawback(s)* occurs twice. Similarly, FIND OUT + *example (s)* returned only one instance, while FIND + *example(s)* has a frequency of 290. In the last two examples [7-55]-[7-56], the Chinese learners confuse FIND OUT with FIND when job seeking is involved.

These findings and the earlier ones about specific patterns suggest that Chinese learners have fewer problems with the use of specific structures of FIND and FIND OUT. That is, the Chinese learners do not confuse the patterns that characterise the single verb or the PV. They are not very likely to apply a specific structure belonging to one verb to the other verb. However, the Chinese learners do suffer from the proper semantic associations such as the collocations (e.g. *trend(s)*) or more extensive units (e.g. [place], [human senses], etc.) which are permissible with only one of the two verbs. They easily stumble over these occasions where the near-synonyms look alike (because there are no salient signals like structures which will sound the alarm to the Chinese learners). As such, these semantic association distinctions are not picked up by the Chinese learners during their learning process.

[7-48] That boy only his Grandpa was in [sn8,s-] , [sn9,-] Grandpa Li found out Wanghua near wanghua's [fm3,-] home.

[7-49] On his arrival, the bee-keeper found out the bees near the back wheels and took them home.

[7-50] "Wang hua [fm3,-] hited me in the face!" when Li Ming walk to him, he find out Li Ming's left eye had already swelled. "

[7-51] Crusoe stepped into the cave with a burning stick, and found out that two eyes he had seen was a dead goat's eyes.

[7-52] From the chart, we can easy [wd2,1-2] find out two **trends** about the healthy [cc4,-1] condition in developing countries.

[7-53] You may find out the **drawback** of the system, that was [vp6,s-], the graduates always were not assigned the suitable jobs according to their situation.

[7-54] It is not difficult for us to find out the **examples** around us.

[7-55] At first, they may find out a job which more suitable for themself [fm1,-].

[7-56] I'll find out my favorit [fm1,-] job and devote my life to it.

## 7.5 Summary

This chapter has explored the Chinese learner language features of two PVs, with greater emphasis on the phraseologies in wider environment. The findings for the phrasal verb CARRY OUT evidenced that the Chinese learner language is not like native English in respect of word-form distribution, collocation and semantic

sequences. For one thing, the Chinese learners do not often utilise the common expression ENFORCE + *law(s)*, but prefer CARRY OUT + *law(s)*. For another, important expressions such as [events] + *be carried out* + *on* + [entities] and [actions/activities] + *carried out* are missing in the Chinese learner language. Instead of these salient usages, idiosyncratic usages such as [authorities] + *carry out* + [public affairs] + [consequences] and *carried out* + [results] are widely present in the Chinese learner language.

The findings of the patterns around FIND OUT have also brought to light the similarity and disparity of the Chinese learner language and native English. The fact that all of the four patterns following FIND OUT are covered in the Chinese learner language indicates that NNSs can use FIND OUT in an idiomatic way to a similar extent to the NSs. Although an overall view of the Chinese learner data may lead us to believe that the Chinese learners are competent in accurately using the PV, fine-grained analysis reveals that the NNSs do not acquire this knowledge as fully as the NSs. The evidence clearly demonstrates that the FIND OUT + *wh-word/how* pattern, which is the central pattern of usage in L1 English, is underused by the Chinese learners. The Chinese learners more strongly tend to employ the other structure, the VN pattern in relation to FIND OUT. The lexical collocates in the N

position of the VN structure, and in the [evaluation] position of the other pattern, *it is* [evaluation] *to* FIND OUT were also examined. The NNSs prefer to use certain noun types which are not frequently used in L1 English with the VN structure, and also repeatedly employ one word, *possible*, in the [evaluation] slot. These features all come to reveal that the Chinese learner language does not completely conform to NS English and lacks the extent of variability exhibited by NSs.

A further analysis comparing the synonymous FIND and FIND OUT was also performed. The patterns of FIND and FIND OUT were identified first in general English. This shows, on the one hand, that in many cases FIND and FIND OUT overlap to a certain extent, or it can be said that there is only a slightly different meaning change, as found in the *V that* structure and the collocation. In this situation, the two verbs are allowed to interchange freely, so they will not be presented as problems in the Chinese learner language. However, on the other hand, there are areas where the usages with one verb are incompatible with the other verb. Pointing out what usage is permitted with only one but not the other verb helps to clarify the dissimilarities between the two verbs. I studied the specific usages which characterise FIND, and discovered that the Chinese learners do not mistakenly apply the phrasal verb to the places where FIND is more appropriate. However, scouring the learner

concordance lines reveals that the Chinese learners confuse the two verbs at times, as signified by the improper collocates or semantic associations. Therefore, an important finding from comparing the two near-synonyms is that learners are less likely to confuse usages which have specific structures (e.g. FIND + Adj, FIND + *no*), however in those circumstances where the structures are the same (e.g. VN, V *that*), but there are semantic restrictions (e.g. collocation with *job(s)*, *trend(s)*, *drawback(s)*, *example(s)*, and semantic fields with [places], [human senses]), learners are apt to mix up the two verbs.

The above findings all suggest that learner language suffers from mis-combination and imprecision of phraseology. In the next chapter, I will further analyse these characteristics and extend the study to cover PVs with other particles.

# **Chapter8: PHRASAL VERBS WITH PARTICLES ‘ON’, ‘ABOUT’ AND ‘DOWN’**

## **8.1 Introduction**

In Chapter 6, I touched on the phraseological patterns in the environments of PVs with the particle UP. Chapter 7 took a step further to lay more emphasis on the patterns which arise from the investigation of PVs with OUT. In this chapter, I extend my research to cover PVs with the particles ON, ABOUT and DOWN.

The purpose of this present chapter is to look into the phraseological patterns of more varieties of PVs. The previous two chapters have demonstrated that the phraseologies around a phrasal verb can reveal learner-specific usages. Finding out whether this can also be attested in other phrasal verb groups is crucial, since it will enhance the validity of this research. The research questions are:

- In terms of distribution, what are the frequencies of PVs with ON, ABOUT and DOWN? What are the most frequent PV types in CLEC and LOCNESS? What are the type-token ratios?
- How do the phraseological patterns (in the groups of ON, ABOUT and DOWN) in CLEC differ from those in LOCNESS?

In order to identify PVs with each of the particles, some steps are necessary to be taken. As reported in Chapter 5, the corpora were annotated with CLAWS tags, which need to be removed in order for WordSmith4 to obtain accurate results. After detagging, the data of relevant PVs was identified using the 'Pattern' facility in WordSmith4. The 'Pattern' facility will show the word which occurs most frequently in a position (e.g. L1 means the first position on the left side of the node word). To obtain frequent patterns, only the words occurring in the L1-L3 and R1-R3 positions and the first four rows are examined. The minimum occurrence threshold is set to 2, so words which occur less than two times will not be shown. By doing so, the most frequent words in each position can be revealed and it is hoped some patterns can be identified as well. The patterns extracted by WordSmith4 will be presented in the figures in each sub-section (see Figure 8.1 and Figure 8.2 in Section 8.2.1 for the display of the 'Pattern' function of WordSmith4).

In the case of the ON group, a total of 357 and 152 lines were captured from CLEC and LOCNESS respectively after cleaning up the irrelevant instances. The frequencies of the ABOUT group are 91 in CLEC and 44 in LOCNESS. There are 447 and 112 hits of PVs with DOWN retrieved from CLEC and LOCNESS. The detailed frequencies and verb types of these three groups are displayed in Appendix C.

The type-token ratios of these three particles groups are 9.5%, 20.9%, 14.8% in CLEC and 19.7%, 20.5%, 37.5% in LOCNESS, for the ON, ABOUT, DOWN groups respectively. The next section reports on the results of the Verb + ON group, followed by a section dealing with the Verb + ABOUT group, and Section 8.4 presents the findings for the Verb + DOWN group.

The frequency of verbs does not guarantee there are sufficient occurrences in both corpora. For example, we will see later that although the most frequent verb KNOCK DOWN is frequent in CLEC, this phrasal verb occurs only once per million words in LOCNESS. Hence, a better method is to select targets from candidates which have sufficient frequencies in both corpora. A cut-off point is not set because it is unnecessary as long as the candidate targets are selected from the very frequent verbs. Two phrasal verb types are chosen from those which occur sufficiently frequently in both of the two corpora. The PVs which occur in both corpora with roughly equal frequencies will be selected first. That is, the difference in absolute frequencies across the corpora is small. This seems to be more justifiable in interpreting the qualitative results with two data sets of similar frequencies.

## **8.2 PVs with ‘ON’**

The summary of the frequency table of Verb + ON (see Appendix C) shows that the

most frequent phrasal verb found in both corpora is GO ON, which occurs 114 times (i.e. 106 per million words) in CLEC and 55 times (170 pmw) in LOCNESS. GET ON (41 times=38 pmw) and KEEP ON (41 times=38 pmw) are also considered frequent in CLEC, but they only occur 5 and 1 times respectively in LOCNESS. The most frequent PV next to GO ON in LOCNESS is TAKE ON (33 times=102 pmw). In Section 8.2.1, the most frequent PV, GO ON, will be analysed first, followed by TAKE ON which is frequent in LOCNESS (33 times) and also occurs in sufficient numbers in CLEC (31 times).

### **8.2.1 GO ON**

The most frequent collocates of GO ON identified by WordSmith4 are displayed in Figure 8.1 and Figure 8.2. Examining the words yielded by WordSmith4, it is clear that not all of them constitute complete units with the phrasal verb. Unfortunately the software cannot highlight meaningful patterns automatically, so I have to look at Figure 8.1 and Figure 8.2 carefully to discover potential interesting phrases. The words which appear to be neither specifically relevant to the node phrasal verb nor constructing any interesting sequence (e.g. *the* + GO ON, *is* + GO ON) were not considered further. The procedure of identifying phrases is to consider the collocates which form meaningful units with the PV from the first to the fourth row. For

example, for CLEC, ‘*time* + GO ON’ was first identified from Figure 8.1, and then GO ON *with*, and so forth. As said, a combination such as ‘*the* + GO ON’ is not considered interesting as the collocate *the* is not directly related to the PV but more related to the noun preceding the PV (e.g. “*as the situation goes on*” in [CLEC]). The judgement of the phrases being interesting targets or not may be argued to be somewhat subjective, but their occurrences were then confirmed through checking all the concordance lines to ascertain these phrases not only actually exist in the concordance lines but they also occur relatively frequently as evidenced by the percentages. The identified patterns were ranked according to their frequencies and percentages, which were recorded in Table 8.1.

N	L3	L2	L1	Centre	R1	R2	R3
1	THE	THE	TIME	GO	ON	WITH	THE
2	HOW	IF	TO	GOING		AND	ON
3	OF	AS	IS	WENT		IN	THIS
4	AND	WHAT	AND	GOES		THE	YOUR

**Figure 8.1: Patterns of GO ON captured by WordSmith4 in CLEC**

N	L3	L2	L1	Centre	R1	R2	R3
1	THE	THE	THEN	GOES	ON	TO	THE
2	THIS	STATES	WHAT	GO		IN	SAY
3	HE	WHAT	TO	GOING		FOR	STATE
4	THEY	OF	HE	WENT		AND	ON

**Figure 8.2: Patterns of GO ON captured by WordSmith4 in LOCNESS**

**Table 8.1: Patterns of GO ON**

CLEC			LOCNESS		
Pattern	Freq.	%	Pattern	Freq.	%
	(Total=114)			(Total=55)	
<i>(As) time GO ON</i>	20	17.5	<i>GO ON to (say, state...)</i>	22	40
<i>GO ON with</i>	13	11.4	<i>GO ON in</i>	8	14.5
<i>to GO ON</i>	10	8.8	<i>then GO ON</i>	6	10.9
<i>GO ON in</i>	8	7	<i>GO ON for</i>	5	9
<i>if N GO ON</i>	6	5.3	<i>what goes on</i>	4	7.2
<i>and GO ON</i>	6	5.3	<i>to GO ON</i>	4	7.2
<i>what is going on</i>	5	4.4	<i>and GO ON</i>	3	5.4
<i>GO ON and ON</i>	4	3.5	<i>(As) time GO ON</i>	1	1.8
<i>GO ON like this</i>	3	2.6			

The frequencies of GO ON in CLEC and LOCNESS are 114 and 55 (see Table 8.1); the proportion of each pattern is calculated based on these numbers. Note that some shorter sequences may be parts of a longer one (e.g. *GO ON like this* co-occurs with *if..GO ON*); thus the cumulative frequencies may not equal to the total frequencies. Patterns which only occur in one corpus are shaded in grey.

As the table apparently shown, parts of the set phrases are common in the two corpora and parts are different. The same phrases include *(as) time GO ON*, *to GO ON* and *GO ON in*, etc., and the different phrases which only occur in CLEC include, e.g. *GO ON with*, *if N GO ON*, while the native-only phrases are *GO ON to (say, state...)* and *then GO ON*, etc.

The pattern *(as) time GO ON* (also including two examples of *days/years GO ON*) is identical in form in both corpora, but is frequently used by the Chinese learners (17.5%) while occurs only once in the native data (1.8%). The reason of such a disparity is probably relating to the genre types of the two corpora, as the nature of essay writing (the major genre type of LOCNESS) seldom involves describing time progression, and the anecdotal style of the Chinese learner writings undoubtedly requires more chances of this expression (see Figure 8.3). Since the numbers of occurrences are not a native-vs.-nonnative difference, I looked into the learner instances to identify aspects of phraseology, in order to see whether the usages in both corpora are similar. Because there is only one occurrence in the native corpus, the BoE was consulted to reveal the main types of this pattern in English. The BoE yielded a total of 606 instances of *(as) time GO ON*<sup>vii</sup>, from which two major semantic sequences were discovered:

- [Previous situation] + *as time* GO ON + [change] + [affect of the change]
  - [old situation] + *as time* GO ON+ *become/start* + [new situation]
  - *as time* GO ON+ *have/has* (perfect)
  
- *as time* GO ON + [increase of degrees]

Both of these two types can also be found in the Chinese learner language as shown in Figure 8.3. The first type can be further categorised into two subtypes. The first subtype indicates a new event which contrasts with the old situation, usually signalled by verbs such as BECOME, BEGIN, START, as exemplified by citations [8-1]-[8-2] in Figure 8.3. The second subtype refers to the completion of an action, usually represented by the past tense or perfect as in [8-3]-[8-4]. The semantic sequences which illustrate the aggravation are demonstrated by citations [8-5]-[8-6]. Probing the uses of these fixed phrases reveals that the Chinese learners have no problem with them.

[8-1] At first, I couldn't grasp [cc3,2-2]the spirit on [pp1,-]how to go [cc3,2-3]over my knowledge, so I was pulled [wd3,-]by them. **As time went on**, I started to master [cc3,3-2] some ideas and I felt I was going into the state.

[8-2] people invented money, just for the sake of convenience. And it really did. **As time went on**, money became a kind of symbol of wealth....Then, too much difference between the rich and the poor made a society unstable.

[8-3] there is a long way to go When euthanasia is legalized, ...**As time went on**, many countries have legalized mercy killing, ..., euthanasia should also be legalized in China.

[8-4] When I was little [wd3,-]... So I wanted to be a judge or lawyer. **As time went on**, my thought had changed [vp6,-]...Then I want to be a manager.

[8-5] **As time went on**, my interest in football became more and more.

[8-6] **As time went on**, you'll become older and older.

**Figure 8.3: Instances of 'as time GO ON' in CLEC**

Another common pattern *to go on* occurring in both corpora shows a slight discrepancy between the two language varieties. Two subtypes of sequences are perceived in the native language: *to go on* + [life] (see [8-7]-[8-8] below) and *to go on to* + [higher education] (citations [8-9]-[8-10] below). In contrast, the Chinese learner language shows two different major subtypes: *to go on with* + N and *to go on* + Ving (citations [8-11]-[8-20] below), both patterns occurring 40% (4 out of 10 cases).

[8-7] Violet and Joe cope with the tragedies of their past and present, and are somehow able **to go on**. [L]

[8-8] Egisthe has had enough of life and no longer wants **to go on**. [L]

[8-9] Most of the pupils leaving the Lycée technique will go to work, although they can take other exams if they wish **to go on to** higher education. [L]

[8-10] Yet amongst the job losses through this, in the UK students in sixth form wishing **to go on to** university are nearing the highest level they've ever been. [L]

[8-11] ...things will in a mess and you can hardly decide how **to go on** . [C]

[8-12] I 'm going **to go on** to learn it when I finish learning volleyball ..[C]

[8-13] need more fresh water ,, the factories need more water **to go on with their work** ,, and lots of fresh water has been...[C]

[8-14] vocation of 1992 ,, I got the notice that asked me **to go on with my study** at Sun Yat-sen Medical Academy ..[C]

[8-15] Some people like **to go on with one job** all through .. [C]

[8-16] So they do n't adjust himself **to go on with a new job** .. [C]

[8-17] and living necessities He or she will be healthy enough **to go on working** .. [C]

[8-18] There he met with Engels who supported him **to go on writing** articles .. [C]

[8-19] How can people allow the polluting factories **to go on destroying** the nature ?? [C]

[8-20] I must admit my English is too terrible for me **to go on studying** it .. [C]

Although the Chinese learners do not combine *to go on to* with [higher education], they were found to employ *go to* with *university*, as the following instances show. The Chinese learners do not make errors using *go to university* (examples [8-21]-[8-26]) but their English can be improved with the advanced use of

‘GO ON *to university*’, which denotes more precise and rich meaning than ‘*go to university*’, a simple phrase usually instructed to the students at a beginning stage.

[8-21] I hoped I could [vp9,3-6] go [vp1,7-4] to university after three years

[8-22] omal [fm1,-] middle school and was entitled to go to University [fm3,-] .

[8-23] al schools and take some courses, part of them go to university after high school to get higher education.

[8-24] y post [wd7, 1-] have governed her thought: going to university, finding a good job and earning a lo

[8-25] She want [vp3, 1-] to go to university because she wants to gain more knowledge

[8-26] standard is not so high, and fewer students can go to university comparing [vp5,s-] to...

One minor but interesting phenomenon is worth mentioning. In LOCNESS, it is noted all of the instances including the pattern *to go on* contain a group of ‘modal-like’ verbs such as *want, wish, be able* (citations [8-7]-[8-10] above). This is also corroborated by consulting BoE: the co-occurring verbs preceding this pattern include volition verbs (e.g. *want, be going to, decided, prepared, like*), permission verbs (e.g. *allowed*), obligation verbs (e.g. *have to, expected, need to, got to*), ability verbs (e.g. *be able to*), and likelihood verbs (e.g. *seemed, likely*), etc. A few such usages can also

be found in the Chinese learner data. For example, the instance *I am going to go on* [8-12] and *like to go on* [8-15] display volition; other modal-like phrases such as *need to go on* [8-13], *asked to go on* [8-14], and *allow to go on* [8-19] convey obligation and permission. It seems the Chinese learners can also use this pattern well, though not to the same extent as the NSs. There may be a semantic/logic reason behind this connection of ‘modal-like verbs’ and ‘to go on’; for example, descriptions of difficulties should often be predicted in the previous discourse so that the person themselves would make a decision to go on or others would encourage or persuade them to continue. A modern Chinese balanced corpus was consulted<sup>viii</sup> and evidenced that the concepts involved are similar in both English and Chinese, i.e. not language specific. In the areas where the first and second/foreign languages are congruent, the Chinese learners may face fewer difficulties.

As to the patterns occurring in only one corpus, the most frequent sequence in LOCNESS is GO ON *to*. This pattern is often followed by a verb, i.e. GO ON *to* + V (68% of the instances containing GO ON *to*), and the interest lies in that the verbs in this slot are mainly speech verbs such as *say, state, tell, explain* and *argue*, accounting for 80% of ‘GO ON *to* + V’. There are only 3 instances (*write, study* and *learn*) of this pattern found in CLEC and none of these verbs are speech verbs. The abundant

use of speech verbs could be a result of the more formal and academic nature of the native corpus. Among the articles, in many places the speech verbs function to emphasise the important parts of the opinions from one specific person. As this pattern is unavailable in CLEC, an alternative approach is adopted. If we find out the environments where one of the speech verbs and a similar meaning of 'GO ON *to*' are present together, it will show that the rarity of 'GO ON *to* + V' in the learner language cannot be explained as simply lacking chance using speech verbs. To investigate this, I searched for the five frequent speech verbs (*say, state, tell, explain, argue*) in CLEC and probe the full sentences to look for instances expressing similar meaning as 'GO ON *to* + V'. One instance "*the boss continued to explain*" was found in CLEC. Similar phenomena can also be observed from another interesting pattern GO ON *for* [time] (e.g. *days, months, years, centuries*), which is absent in CLEC. I queried CLEC to discover which verbs precede *for days/months/years/centuries* and collected two verbs LAST, KEEP ON, which have nearly equivalent meaning as GO ON.

[8-27] At the beginning, I tried to write one short passage everyday, I kept it on for months.[C]

[8-28] The Spring Festival usually lasts for 20 days. [C]

It is also suspected in passing that the Chinese learners use other syntactic

structures on occasions where a time phrase with *for* is more appropriate or expressive. The concordance lines of GO ON in CLEC were probed carefully to look for such uses. One example was obtained: “ *Although smokers can not feel anything in a shore[sic] time, its effects are going on continuously*”, in which the adverb *continuously* can be revised as *for a long time/years*'.

It is dangerous to jump to the conclusion that the Chinese learners do not know the words or phrases given the lack of certain pieces of data in the Chinese learner corpus. In fact, by examining the paradigmatic variations of the collocation slots, the findings ascertain that the Chinese learners employed alternatives in place of GO ON in the patterns which were thought inaccessible to them. Corpus data cannot tell us whether the writer has the knowledge about GO ON or not at the time he produced “*continued to explain*”, “*lasts for 20 days*”, etc. What we can be certain is the Chinese learners do have opportunities to use ‘GO ON + Speech verbs’ and ‘GO ON *for* [time]’, but they opted for other choices. There is nothing wrong in the Chinese learners choosing synonymous words or phrases to replace GO ON and *for* [time]. However, this learner performance indicates the need to highlight the two patterns ‘GO ON + Speech verbs’ and ‘GO ON *for* [time]’ in our EFL courses with the purpose to increase their rhetoric variations.

Another contrast that attracts my attention is that the Chinese learners tend to use the phrase *what is going on* while the NSs prefer *what goes on*. Their concordance is shown in Figure 8.4 below. Citations [8-29]-[8-33] are for *what is going on*, which occurs in CLEC, and they mainly follow a prepositional phrase such as *inform of*, *focus on*, etc. However, in the native language, even when a preposition phrase is present, as in citations [8-35]-[8-36], the simple tense *what goes on* (which implies more generality) rather than the progressive tense (which emphasises the present situation) is employed by the native students. *What is going on* is more related to concurrent events while *what goes on* is more related to hypothetical events or general states of affairs. Again, this contrast may be largely affected by the styles of the two corpora: one comprises mostly narrations and the other essays. In addition to that, other factors might also have effects. All the examples seem to follow this tendency, except example [8-31], which criticises housewives for being ignorant. This novice Chinese author is not referring to any immediate events but general or new information. Apparently this case demonstrates such a distinction is not crystal clear to some Chinese learners. A possible explanation may be the Chinese-English difference: a time indicator 'now' is always required with the expression *what is going on* in the Chinese language but English differentiates the two by aspect. Without the

time marker, the distinction between these two phrases is blurred to the Chinese.

[8-29] The program keeps the masses well **informed of what is going on** in the world. [C]

[8-30] Outside the stadium, patriotism draws people to TV, newspaper, **focusing on what's going on** in the games. [C]

[8-31] She will have **no idea of what's going on** except the price in the market.

[8-32] You've mentioned in your letter that you will be back today, right? What's going on?" "It's too bad. I can't get the ticket.[C]

[8-33] To keep us **informed of what is going on** home and abroad. [C]

[8-34] Therefore, it would be difficult to control what goes on in the barracks. [L]

[8-35] This will mean that everyone will have a fair say **in what goes on**-something which must be guaranteed for...[L]

[8-36] ...it has enabled us to widen our understanding **of what goes on** in the world by allowing us access to...[L]

[8-37] ...that would result from showing the public what goes on behind the closed doors of the execution...[L]

**Figure 8.4: Examples of 'what is going on' and 'what goes on'**

In this survey of patterns of GO ON, we have seen cases which display large discrepancy in frequencies between the corpora but are employed qualitatively by the Chinese learners as well as the native students (e.g. *as time* GO ON, modal verbs + *to go on*), and also cases which occur about the same times in both corpora but show contrast in the major pattern varieties (e.g. *to* GO ON). For those which only found in the native corpus (e.g. GO ON *to* + [higher education]/Speech verbs), the substitutes are recognised by identifying similar occasions in CLEC. In search of the replacements, evidences such as *go to university*, CONTINUE, LAST, KEEP ON are collected, corroborating the Chinese learners are not short of opportunities making use of the phrasal verb; instead, they settled on other options. Other findings such as the variances of ‘*what* GO ON’ highlight the Chinese learner language preference which may be attributed to their L1.

### **8.2.2 TAKE ON**

We now move on to the other frequent phrasal verb, TAKE ON. The same procedures are applied to this PV. The collocates in the positions from L3 to R3 are presented in

Figure 8.5 and Figure 8.6, and the patterns identified from these collocates are listed in Table 8.2.

N	L3	L2	L1	Centre	R1	R2	R3
1	PEOPLE	LIKE	TO	TAKE	ON	A	NEW
2	THE	I	WILL	TOOK		THE	KIND
3	T	NTRIES	HAS	TAKING		ONE	SAME
4	ARE		ALWAYS	TAKEN			ALL

**Figure 8.5: Example patterns of TAKE ON in CLEC**

N	L3	L2	L1	Centre	R1	R2	R3
1	THE	TO	TO	TAKE	ON	THE	FORMS
2	BUT	AS	AND	TAKEN		A	SINS
3			HE	TAKES			FORM
4			HAS	TOOK			THAT

**Figure 8.6: Example patterns of TAKE ON in LOCNESS**

**Table 8.2: Patterns of TAKE ON**

CLEC			LOCNESS		
Pattern	Freq.	%	Pattern	Freq.	%
		(Total=31)			(Total=33)
TAKE ON <i>a</i>	18	58	TAKE ON <i>the</i>	16	48.5
TAKE ON <i>a</i> ( <i>new/peaceful</i> <i>look</i> )	10	32	TAKE ON...FORM	6	18
TAKE ON... <i>job</i>	9	29	TAKE ON ( <i>sins/guilt</i> )	6	18

Table 8.2 shows the two corpora have a common pattern ‘TAKE ON + *look/form(s)*’, though the Chinese learners use *look* only and the native preference is *form(s)*. Either *look* or *form* makes similar senses with TAKE ON denoting ‘develops a new appearance or quality’, but there is fine difference since the former suggests refreshing of something and the latter indicates transforming to a particular shape, as in “*Everything takes on a new look as Spring Festival comes*” and “*He takes on the form of a god demanding deaths*” from CLEC and LOCNESS. The phrase ‘TAKE ON + *a* +Adj +*look/job*’ accounts for 61% of the Chinese learner data, and the same patterns appear to recur as exemplified below (Figure 8.7).

[8-38] n job-hopping is that I can always take on a kind of job if I can appl

[8-39] Some people inclined [vp1,-2] to take on a kind of work from beginni

[8-40] ever. Some are [vp9,-1] intend to take on a kind of job throughout th

[8-41] do [wd5,s-] the people around you take on a new look everyday [fm2,s-

[8-42] oolhouse in it. Now our school has taken on a new look. There are one

[8-43] o schools are also built. The town takes on a new look so I love the m

[8-44] technology to us. Our country has taken on a new look since we had [c

[8-45] middle school. The study problems take on a new look before us. So fi

[8-46] ng of their houses. Everything takes on a new look as [np7,-2] Sp

**Figure 8.7: Examples of ‘TAKE ON + a +Adj +look/job’**

It seems that the Chinese learners repeat the two patterns ‘TAKE ON *a/an* ADJ *job*’ and ‘TAKE ON *a/an* ADJ *look*’ time and again. This could be topic-induced because the same nouns recurred, but a look into the text sources shows that the patterns occur across different texts and were employed by different authors. This

implies there is particular consistency in the Chinese learner language, which cannot be explained in light of that the Chinese learners are writing or talking about the same things. The possible reasons left may be attributed to instruction input or L1-L2 difference.

Apart from the major patterns mentioned, a contrast was also found in the remainder of the instances. The Chinese learners tend to connect TAKE ON with undesired events, but such a negative connotation was less detectable in LOCNESS. As demonstrated by Figure 8.8 and Figure 8.9, the discourse of the Chinese learners' data often suggests the difficulty or danger of the task, but the NSs tend to describe the qualities or properties of what they take on, thus are neutral.

- what is **the opportunity that most people do n't like** to take on ,, above all that ,, there are some way that we c 20
- they want to live a stable lives and do n't want to take on **danger** ..
- opinion ,, however **bad or good work** it is ,, I will take on it all time .. 23
- As a wife and a mother ,, she had to take on such **heavy and time-consuming housework** that she had no
- while for the second kinds of people ,, they usually take on the **danger** of finding their job .. 31
- of peoples they are suitable and stable ,, they need n't take on the **danger** of not finding suitable job .. 30

**Figure 8.8: Concordance lines of TAKE ON in CLEC**

- system If in the last years of his presidency ,, d'Estaing took on a **more prominent role** then ,, it was only partly
- the South decided to become independent ,, they decided to take on a **certain way of life** ..
- begun to increase before the end of the war ,, as people took on a **care-free attitude** ,, with little feeling of responsibilities 35 32
- for the future ,, the people fall back into the past and take on the **value of an object** -- 'tre en soi' ..
- He refuses to take on **the values and traditions** of his own town ,
- he rejects the fact that he can make decisions and takes on **the state of an object** ..

**Figure 8.9: Concordance lines of TAKE ON in LOCNESS**

The results of TAKE ON show the Chinese learners sometimes are prone to the use of simple and fixed phrases such as ‘TAKE ON +a +Adj +look/job’. The inclination towards such basic patterns is in keeping with reports such as Kamakura (2007:10), supporting that ‘repetitiveness’ is an ascertained feature of learner language (Ringbom, 1998:50; Milton, 2000:14). The contrast of semantic prosody has also been observed by Kamakura (2007:19) in his example of *in the world*. Such contrast is also witnessed in the case of TAKE ON, which may be characteristics of the Chinese learner language as well. The learners’ preference of taking on negative events is possibly affected by their L1, because TAKE ON is often translated to ‘undertake challenges’ in Chinese.

### 8.3 PVs with ‘ABOUT’

The second table in Appendix C lists the frequencies and verb type of PVs with ABOUT, from which BRING ABOUT was found to be the most frequent PV in both corpora, thus will be examined first.

#### 8.3.1 BRING ABOUT

Before the patterns of BRING ABOUT are examined, it is noticed that the noun collocates of BRING ABOUT in CLEC reveals some instances of which the nouns are not appropriate collocations, suggesting some learners do not have full knowledge in relation to this PV. These instances include: *fact*, *convenience*, *inconvenience* and *knowledge*. None of these instances are found to follow BRING ABOUT in BoE. The patterns of this PV identified by WordSmith are given in Figure 8.10 and Figure 8.11, from which meaningful patterns are collected and summarised in Table 8.3.

L3	L2	L1	Centre	R1	R2	R3
THE	IT	WILL	BRING	ABOUT	THE	A
AND	KILLING	HAVE	BROUGHT		OF	LOT
MERCY	WHAT	MEANS	BRINGS		A	PROBLEMS
	TO	HAS	BRINGING		GREAT	TO

**Figure 8.10: Patterns of BRING ABOUT in CLEC**

L3	L2	L1	Centre	R1	R2	R3
OF	AND	WOULD	BROUGHT	ABOUT	A	THE
		TO	BRING		THE	BETTER
		AND	BRINGS		BY	
		IT				

**Figure 8.11: Patterns of BRING ABOUT in LOCNESS**

**Table 8.3: Patterns of BRING ABOUT**

CLEC			LOCNESS		
Pattern	Freq.	% (Total=59)	Pattern	Freq.	% (Total=27)
BRING ABOUT <i>the</i>	12	20.3	BRING ABOUT <i>a/an</i>	7	26
<i>bringing about of</i>	8	13.6	BRING ABOUT <i>the</i>	4	14.8
<i>have(has) brought about</i>	8	13.6	<i>would bring about</i>	4	14.8
BRING ABOUT <i>a/an</i> <sup>ix</sup>	5	8.5	<i>brought about by</i>	3	11.1
<i>will bring about</i>	5	8.5	<i>it</i> BRING ABOUT	2	7.4
BRING ABOUT <i>great</i>	3	5.1			

From Table 8.3, we could see the most frequent pattern in both corpora can be transcribed as ‘BRING ABOUT + determiner + N’. However, the NNSs and NSs differ in the semantic prosody this sequence show, as suggested by the noun collocates. The negative connotation is much more strongly expressed in CLEC,

whereas the native corpus LOCNESS shows more neutral meanings. 61% of collocates in CLEC hold a negative implication (e.g. *problem, mistake, failure, disadvantage, difficulty, loss, trouble*), or are modified by a negative adjective (e.g. *bad reaction, ill effect*). Most of the collocates in LOCNESS do not display this tendency, the examples being *awareness, population, life, attitude, society, change, result, improvement, trend, justice, revolution, recognition, productivity, etc.*

Examining the most frequent nouns on the right-hand side of the node PV in BoE also shows collocates such as: *change, peace, downfall, release, collapse<sup>x</sup>, death, improvement, revolution, reconciliation* (in descending order and lemma form). This finding chimes with the study of the same PV by Johns (1997), where his data on BRING ABOUT exhibits a preponderance of positive connotations, as well as Xiao and McEnery (2006:115). However when *problem, mistake, failure, etc.* are talked about in the learner data, a negative feeling is obtained inevitably. Five examples were taken every 10<sup>th</sup> line from the concordance lines of this PV in CLEC, as exemplified by citations [8-47]-[8-51] below, suggesting the Chinese learners are inclined to link BRING ABOUT with negative consequences.

[8-47] ...of serious unemployments, it brings about other problems at the same time .. [C]

[8-48] Haste usually brings about a failure.[C]

[8-49] The incident brought about immediate oppositions and blames of the international community... [C]

[8-50] Therefore , it brings about some mistakes in examine &...[C]

[8-51] Second, many fake commodities can bring about many difficulties to the factories which make the real commodities. [C]

Besides the central patterns in relation to the phrasal verb BRING ABOUT, there are also some patterns which are actually marginal in English but are over-represented in CLEC. Twenty-nine per cent of the concordance lines of BRING ABOUT in CLEC are found to co-occur with phrases which emphasise the amount (e.g. *a series of, a lot of, a great deal of, countless, great, many*), or adjectives which denote an increase or intensification of what was brought about (e.g. *better, more and more*). However, this is not observed in LOCNESS, where only 4 cases (14%) can be said to follow this pattern. To ensure that the semantic sequence '[quantity/intensification] +N' is not a typical pattern for BRING ABOUT, 100 instances randomly selected from BoE were closely investigated. Among them, only 5% of cases have this semantic sequence,

confirming that the Chinese learners appear to overemphasise the importance of this sequence.

### 8.3.2 COME ABOUT

After examining BRING ABOUT, there are no other verbs which have sufficient frequency for further analysis. However, one of the verbs seems to be worthy of some discussion here: the phrasal verb COME ABOUT. This PV is studied by Partington (2004) in comparison with HAPPEN. He finds that HAPPEN tends to co-occur with ‘unfavourable/unpleasant’ things, or indicates something occurs ‘by chance’ or expresses the ‘lack of factuality’. He also observes that COME ABOUT occurs in company with words emphasising ‘process’. However, this is less evident in the present corpora, since its occurrence number is small in the native corpus.

This phrasal verb COME ABOUT is not found in CLEC but occurs 9 times (28 times pmw) in LOCNESS. In addition, this PV has similar meanings in relation to BRING ABOUT, albeit with different transitivity. Another salient semantic sequence is uncovered by searching the phrasal verb COME ABOUT in BoE: ‘COME ABOUT + [reason/cause]’. The most frequent words which follow the PV are *because (of)*, *as a result of*, *through*, *by (virtue of)*, *after*, *from*, etc., as exemplified below. This implies that the occurrence of something was inevitable, not planned or beyond control, and

an explanation is usually required in the following text.

[8-52] ...Hitler's tyrannical impulse and its unfortunate consequences came about as a result of his being a mediocre artist...[BoE]

[8-53] ... it came about through his innate shyness.[BoE]

[8-54] And it all came about by sheer luck.[BoE]

[8-55] This remarkable change came about after Ed Roberts gained a success consciousness...[BoE]

This sequence is also the dominant one found in LOCNESS. But unfortunately, as stated earlier, the CLEC data does not include the use of this PV at all. In order to test whether this combination of meanings exists in the Chinese learner language, the synonymous verb, HAPPEN, was examined, too. Three instances listed below were found to be possible occasions where the sequence 'COME ABOUT + [reason/cause]' can fit in and the verb can be replaced by COME ABOUT. This is not to say that these three are misuses; it shows opportunities for the Chinese learners to make their language more expressive and varied, but the Chinese learners cling to the verb which they can employ with confidence or they lack the knowledge about infrequent verbs such as COME ABOUT in contrast to HAPPEN.

[8-56] Sometimes tragedy happened because of the bad effect of TV.[C]

[8-57] This change happens because of some factors.[C]

[8-58] This happened after Amelia's father became a ruined man...[C]

In the study of Verb+ABOUT we have witnessed again the Chinese learner language is characterised by some distinct features. In the case of BRING ABOUT, like TAKE ON, the Chinese learners rely more on nouns creating a negative semantic prosody, which is not shown in native texts. We also observed the seriousness of problems are underlined through adding quantities in the Chinese learner language. Lorenz (1998:59), in his study of adjective intensification, explains the overuse of intensification is a conscious strategy employed by learners to impress the reader. This cannot be testified in this research as it does not tackle all intensifiers but only one PV. The Chinese learners' English writing, again, can be improved if they do not make do with the common verb HAPPEN but employ the phrasal verb COME ABOUT, in the context where a necessity of [reason/cause] is required to explain why something is taking place.

## 8.4 PVs with ‘DOWN’

The third group to be researched is PVs with the particle DOWN. The most frequent PVs in CLEC are KNOCK (65 times=61pmw), SIT (46 times=43pmw) and FALL (45 times=42pmw), and the most frequent ones in LOCNESS are BREAK (16 times=49pmw), GO (9 times=28pmw), CUT (8 times=25pmw) and LOOK (8 times=25pmw).

The particle DOWN is often related to ‘movement from a higher position or level to a lower one’ (Sinclair et al., 2002:461). Apparently, the PVs used frequently in CLEC generally denote physical actions. Even for a PV which may have both literal and idiomatic meanings like *knock down*, a majority of instances in the Chinese learner language are describing the action. In a word, the three frequent PVs in CLEC are either themselves a pure literal PV (e.g. *sit down*) or are mostly used in their literal meaning (e.g. *knock down*, *fall down*). The phrasal verb KNOCK DOWN is mainly employed by the Chinese learners in its literal usage. Forty out of the 65 instances of KNOCK DOWN (62%) in CLEC contain the formulaic phrase ‘KNOCK DOWN *to the ground*’ which describes someone being hit and falling to the ground during a fight. Although a figurative use, ‘be hit by a vehicle’, is also found in CLEC, the large number of literal uses indicate that most Chinese learners are more comfortable with

those senses rather than the idiomatic ones. A similar phenomenon can also be observed in the examples of FALL DOWN. Seventy-six per cent of the 45 citations of FALL DOWN in CLEC are literal uses in the senses of ‘stumble, drop to the ground’, etc.

In contrast, these highly frequent phrasal verbs of CLEC were rarely found in LOCNESS. They occur only one or two times, rendering them inappropriate for further analysis. As a result, the two PVs, BREAK DOWN and CUT DOWN were selected because they have relatively higher occurrences in both corpora. They will be analysed in detail in the following subsections.

### 8.4.1 BREAK DOWN

IS	IT	TO	BREAK	DOWN	WE	MUST
			BREAKS			
			BREAKING			

**Figure 8.12: Patterns of BREAK DOWN in CLEC**

WHEN	THE	BE	BREAK	DOWN	THE
IS	COMPUTERS		BROKEN		OF
			BREAKING		AND

**Figure 8.13: Patterns of BREAK DOWN in LOCNESS**

The pattern displays of BREAK DOWN by WordSmith are shown in Figure 8.12 and Figure 8.13. Because not many patterns can be summarised from these two displays, no table was compiled. I then manually scrutinise the concordance of BREAK DOWN in order to discover potential patterns. Probing the data shows that 50% of the 14 instances of BREAK DOWN in CLEC are used in the sense of ‘stop working, out of order, broken, damaged’, as shown by the following examples (Figure 8.14). The remaining lines also contain three instances of the sense ‘collapse’ such as “... *the industry was almost broken down, but also our society will break down*”, and two instances of the sense ‘to interrupt, break, desist from, discontinue’ such as “*break down your habits, break down the old ideas*”.

```

who willingly installs his spare-time on your car breaking down on the way.
When we buy some electronic device, it may break down after a few weeks.
ago, when I was hurring to go home, my bike suddenly broke down I must get it repaired.
buys fake family facilicians , in case they were to break down, the damage would be terrible.
First, they break down very easily because of the poor quanlity .
so when we use them, they don't work well and easy to break down .

```

**Figure 8.14: Examples of BREAK DOWN from CLEC**

However, the dominant pattern in LOCNESS is found to be absent in CLEC. This pattern was not picked up by the pattern function of the WordSmith program but can be identified by looking through the concordance lines. In a good proportion of all

the occurrences of BREAK DOWN (25%) in LOCNESS, the words BREAK DOWN and *barrier(s)* collocate with each other (see Figure 8.15).

---

The community has an aim to **break down** the economic and political barriers  
There are already many proposals under way for **breaking down** the frontiers between countries and  
on the image of women is the first step in **breaking it down**, so that women are seen for who  
goal for school integration programs is to help **break down** the racial barriers and promote better  
When it's all **broken down** it's nothing but discrimination.  
The **breaking down** of barriers, etc, will be seen as  
All trade barriers will be **broken down** so as to encourage and promote free

**Figure 8.15: Examples of BREAK DOWN in LOCNESS**

Clearly this usage ‘to do away with hedges, obstacles’ or figuratively ‘break/change the wrong beliefs’ is central to BREAK DOWN for native speakers. This is further confirmed by consulting BoE. The word *barrier(s)* is found to be the most frequent direct object of BREAK DOWN and appears 186 times (ranked top5) in the R2 position of BREAK DOWN. I searched *barrier(s)* in CLEC, and found two verbs OVERCOME, CAST OFF were employed by the Chinese learners to replace BREAK DOWN. Again, the lack of ‘BREAK DOWN+ *barrier(s)*’ in the Chinese learner language can be complemented by altering the verbs in these environments.

[8-59] it [pr3,s-].In my view, we can overcome the barrier in the development of our economy.  
[C]

[8-60] lking about our reading habits. The initial barrier between strangers was cast off; the c [C]

The meaning ‘out of order, broken’ does not only receive great attention in the Chinese learner language but it is also employed in the native language. However, there is a difference in the usages across the two corpora. In LOCNESS, it often collocates with *computer, car*, etc. (Figure 8.17), indicating the common events when cars and computers stop functioning in our daily life. However, in CLEC, it does not describe the specific items stopping operating, but often underlines the poor quality which leads to the high possibility of damage (Figure 8.16). This shows clearly that the native students often talk about the situation where a breakdown happens, which is evidenced by the frequent collocation with *when* and *if* in Figure 8.17. The Chinese learners, however, emphasise the high possibility that something may fail by stressing their fragility (e.g. *easily*) or a short period of time (e.g. *a few weeks*) or their poor quality (e.g. *fake*). This exhibits the Chinese learners’ different favouring of one usage which is not manifested in the native English.

First, they **break down** very easily because of the poor quality .  
 so when we use them, they don't work well and easy to **break down** .  
 When we buy some electronic device, it may **break down** after a few weeks.  
 who willingly installs his spare-time on your car **breaking down** on the way.  
 buys fake family faciliicians , in case they were to **break down**, the damage would be terrible.

**Figure 8.16: Examples of BREAK DOWN with ‘items of bad quality’ in CLEC**

Also, computers can break down, and then a human is needed to take very dependent on computers when computers do break down.

When computers break down people are thrown back to the old way in case of an emergency; if they are stranded or broken down. bumping can damage the cars and when the cars break down do not blame us, because the cars are you bump other cars, a car that is hit might break down in the middle of the track and can

**Figure 8.17: Examples of BREAK DOWN with ‘computer/car’ in LOCNESS**

### 8.4.2 CUT DOWN

The phrasal verb dealt with in this subsection is CUT DOWN, whose patterns captured by WordSmith are shown in Figure 8.18 and Figure 8.19.

THAT TREES TO CUT DOWN THE OF  
OF I THEY CUTTING TREES  
CUTS SOME  
TO

**Figure 8.18: Examples of CUT DOWN in CLEC**

TO CUT DOWN ON

**Figure 8.19: Examples of CUT DOWN in LOCNESS**

The most explicit pattern in CLEC is the collocation with *tree(s)*. In total, 6 out of the 24 examples of CUT DOWN in CLEC are combined with *tree(s)*, but if all the cases in the sense of ‘cut through/make something fall down’ are included (other collocates are *forest, branch, rose, etc.*), the number increases to 13 examples (54%).

The Chinese learners apparently incline to employ the literal uses of this PV. The second salient meaning is ‘reduce’, which appear 11 times in CLEC (46%), co-occurring frequently with the word *water* in the meaning of ‘saving water resources’. On the other hand, LOCNESS shows one very consistent and special pattern ‘CUT DOWN + *on*’ (50%), which is absent in CLEC. This pattern is utilised in the sense of reducing the amount of something undesired. This pattern is also found to be the most frequent one in BoE (ranked top); the examples cited from LOCNESS are displayed in Figure 8.20.

which can then be used in combustion engines which will cut down on toxic gases which are given out.  
 Cutting down on national speed limit may also encourage  
 that the only way to accomplish that goal is to cut down on have sex before marriage.  
 arguing over how to Reform the Welfare System, cut down on crime, and balance the budget.

**Figure 8.20: Concordance of CUT DOWN ON from LOCNESS**

In order to see whether the Chinese learners choose other verbs to replace this phrase, the frequent collocates in English have to be found. By examining ‘CUT DOWN on’ in BoE, it revealed that this phrase often takes words such as *fat, alcohol, salt, sugar, pollution, caffeine, cigarettes, car, smoking, booze, calories, etc.* (in

descending order) as the direct object. It is usually associated with reducing the amount of an undesired habit which may do harm to our health or the environment. These collocates were examined manually in turn in CLEC. Out of 928 instances, I found two verbs REDUCE and DECREASE, occurring 15 and 3 times respectively, are used in similar occasions with pollution, as shown below (Figure 8.21). Admittedly, the contexts of these examples provide the Chinese students opportunities to make their language more varied and expressive if they do not stick to the same verbs.

*the technology and reduce the pollution. Finally, it is necessary to control the*

*Then we must control the pollution and reduce the pollution. It is important for us to limit the cri*

*of what [fm1,-] the industry uses and reduce the pollution [wd3,s-] to [pp1,-3] the least harm. Thi*

*en water [wd3,s-] the flowers. We must reduce the pollution to protect the resource of fresh water.*

*[sn1,s-] The [wd5,s-] second, we must reduce the pollution to the water, so we can use all the wate*

*not waste water. Second, people should reduce the pollution of fresh water.*

*on. Not only do it, but also we should reduce the pollution. [sn8,s-] To be brief, It [fm3,1-] ought*

*er in the industry. Finally, we should reduce the pollution of water. In short, I believe that we ha*

*h-controll [fm1,-] . Second, we should reduce the pollution. Third, but not last, we should not use*

*nce of fresh water and try his best to reduce the pollution of water.*

**Figure 8.21: KWIC of REDUCE from CLEC**

In this study of Verb+DOWN, we have noted pronounced literal uses of phrasal verbs in CLEC, which may be related the genre difference or the nature of the particle DOWN. We also perceived further evidence supporting previous findings of other PV groups. The Chinese learner language lacks the strings ‘BREAK DOWN + *barrier(s)*’ and ‘CUT DOWN *on*’, which are typical thus important in English. These two sequences are figurative in contrast to the Chinese learners’ preference of literal uses. The Chinese learners, however, were demonstrated to adopt other verbs (OVERCOME, CAST OFF, DECREASE and REDUCE) to express similar meanings. Given the Chinese learners do have to produce similar meanings in their texts, this is a piece of further evidence bearing out the Chinese learners may not be familiar with the sequences or they simply abandon these options. Parts of the results also point to that the LL has consistent fondness of some concepts involved with one particular phrasal verb, which the native language does not have. The connection of BREAK DOWN with [poor quality] of devices in the LL confirmed this.

## **8.5 Summary**

This chapter has presented the results relating to PVs with three particles: ON, ABOUT and DOWN. The frequencies (tokens) and verb types of each phrasal verb

group have been produced (see Appendix C). Furthermore, case studies have been conducted in order to capture the pattern difference (and similarity) between CLEC and LOCNESS. The Chinese learner language's distinctiveness were accounted for by possible sources such as genre types, topics, L1-L2 difference, and instruction input. On many occasions, it is difficult to pin down the reasons to one factor because they may be intertwined and hard to tease the effects apart.

The task of this chapter turned out to be extremely challenging because it suffers from three unavoidable difficulties. The first is that some phraseological patterns have been noticed not occurring in CLEC, a problem that all studies of learner corpora are facing. The lack of data does not necessarily indicate learners' inability, thus I took an alternative approach to discover the occasions where substitute verbs are in use. By doing this, we may reasonably infer that the Chinese learners bypass the use of PVs, regardless whether they know the appropriate PV or not. The second is that the results are fairly individualised, in terms of each phrasal verb and each pattern. The patterns of a phrasal verb appear to be characteristic of that phrasal verb in question, and most of the time are not applicable to others. Furthermore, the features of one pattern are by and large also specific to that pattern. In addition to these two problems, an obvious feature of the findings is their occurrences are not overwhelmingly frequent.

This is an inevitable consequence since the research targets are phrasal verbs, which are not common by nature, let alone my research targets are their phraseological patterns, a minority of the minorities. That each pattern is individual and the small numbers of results render reaching coherent conclusions even difficult. One solution to amend these faults is investigating more PVs, this is why this study has covered six phrasal verb types. These factors are essential to studies of phraseologies in learner languages.

Despite of these limitations, this study of phrasal verbs with the particles ON, ABOUT and DOWN has revealed differences of the Chinese learner language in comparison with native writing. It has been observed that certain patterns show apparent discrepancy in frequency and behaviour differences across LOCNESS and CLEC. For the quantitative difference, some are relatively frequent in one of the two databases only. The discrepancy in numbers does not necessarily indicate usage difference. Sometimes we can witness a few cases (e.g. *as time GO ON*) in which the patterns in the Chinese learner language are not inferior to those in the native variety. For the qualitative difference, some meanings/concepts are found to relate to one particular phrasal verb or pattern by one group of writers but not the other. Through the individualised patterns, the Chinese learner language can be profiled by the

Chinese students' distinctive preferences. Their associations of different semantic prosody or semantic preferences with a PV contrasting the native English suggest the Chinese learners' language has its internal system, which may be largely influenced by their L1 background. For those patterns which do not show in the learner corpus, I explored the occasions where similar meanings are expressed and found alternative expressions were deployed in lieu of the phrasal verbs. Given the low probability of finding these occasions in practice, although the numbers of the instances are not large, they are still evidences demonstrating when opportunities of producing a particular pattern are provided, the Chinese learners may shy away from using it. In the next chapter, the findings of this chapter and those of the previous two chapters will be summarised and discussed in light of literature and reflections.

# **Chapter9: DISCUSSION AND PEDAGOGICAL IMPLICATIONS**

## **9.1 Introduction**

I have compared the behaviours around PVs in a corpus of Chinese learner language (CLEC) and a corpus of native English (LOCNESS) in the preceding Chapters 6-8. In this chapter, I will revisit the research questions and summarise the answers. Section 9.2 addresses the main threads that run through the thesis, that is, the main results. A general discussion of their relationship to the literature and the possible rationales of the findings will be presented in Section 9.3. Section 9.4 is devoted to the pedagogical recommendations that this work may provide to teaching and learning of PVs, and then a synopsis of this chapter is given in the end.

To refresh our memory, the guiding research questions which were presented in Chapter 1 are repeated here:

1. How do the Chinese learners and native English writers use PVs differently in terms of distribution (e.g. frequency of occurrences, type-token ratios and the most frequent PV types)?

2. How do the degrees of idiomaticity and restriction strength help to characterise PVs, based on data from an English reference corpus (BoE)? (Chapter 6)
3. How can a phrasal verb be distinguished from its near-synonym, in the Chinese learner corpus (CLEC) and the English native corpora (LOCNESS and BoE)? (Chapter 7)
4. How do the Chinese learner uses of PVs differ from native uses in terms of phraseological units? (Chapters 6 to 8)

## **9.2 Review of major findings**

This section recapitulates the main points which this thesis has addressed. Each subsection presents answers to a corresponding research question, beginning with a review of the need to answer that question, followed by summaries of the findings.

### **9.2.1 Question 1: The distributional features of PVs**

This thesis probed into the linguistic behaviours of PVs by using corpus approaches. The primary feature that underpins any corpus approach is the frequencies of the target items to be examined, and this research is no exception. The occurrence numbers of all of the PVs can give an overview of phrasal verb distribution, and the high-frequency PVs can lead us to obtain more valuable in-depth analyses. The

frequency data must be recorded before other research questions can be answered. To achieve this research purpose of uncovering the distribution of PVs, I have surveyed the occurrence frequencies of all of the PVs, identified the phrasal verb types (see Appendices A-C), and also compared the type-token ratios. The sums of their cumulative frequencies, the most frequent items across corpora, and the TTRs are synthesised respectively from Table 9.1 to Table 9.3.

First, the occurrences of each group of PVs per million words have been reported in the previous three chapters and are summarised in Table 9.1 below. The Chinese learner corpus CLEC registers more tokens in all of the particle groups, except ON and ABOUT. Although there are group differences, in terms of total frequencies, CLEC appears to contain significantly more uses of PVs overall (chi-square value=75.61,  $df=4$ ,  $p<0.01$ ). The reason for this over-representation of PVs in the Chinese learner language may be a consequence of the Chinese learners employing fewer types but more tokens; in other words, they repeat the same phrasal verb types, which can be corroborated from the TTRs below.

This runs counter to the common hypothesis about PVs in native and non-native writings. Given the bewildering complexities of their syntactic and semantic features (see Chapter 2), PVs are anticipated to have fewer occurrences in learner languages.

This has been confirmed in most cases: for example, McKenny’s (2006: 141) results show that the native corpus (LOCNESS) contains twice more PVs than the Portuguese learner corpus (Porticle). In the same way, Waibel (2007) also documents that in the eleven subcorpora of the ICLE corpus, PVs are under-represented in eight sub-corpora. However, over-representation and roughly equivalent numbers of PVs in three learner languages (German, Dutch, and Polish) are reported in the same study. The varieties of second/foreign language learners of English do not perform identically in respect to quantification. A recent study (Chen, 2013) comparing the quantity of PVs in a Chinese learner corpus with American and British corpora provides support to my work as it also reveals that the Chinese learners do not shy away from using PVs.

**Table 9.1: Normalised frequencies of PVs across five particle groups**

	<b>UP</b>	<b>OUT</b>	<b>ON</b>	<b>ABOUT</b>	<b>DOWN</b>	<b>Total</b>
LOCNESS	1119	1338	469	136	345	3407
CLEC	1521	1497	353	85	418	3874

Second, the TTRs of each particle group are presented in Table 9.2. The total number of types of a group is divided by the total number of tokens of that group, and

converts to percentages. LOCNESS generally has a much higher ratio across the five groups of PVs, that is, the native writers use a wider range of PVs but fewer tokens per type (apart from one group, the ABOUT group, which has similar ratios in both corpora). This is hardly surprising, since the native writers are presumably more proficient and skilful. Nevertheless, the diversity of types and tokens in a learner corpus is not always small. For example, the German and Italian sub-corpora of ICLE have similar TTRs to those of LOCNESS (Waibel, 2007: 86). Considering that the CLEC is much larger than LOCNESS in size, we may expect to find more types of PVs in CLEC but in fact we found fewer. Therefore if the Chinese learners aspire to catch up with native students, they will need to be able to employ more varieties of phrasal verb types.

**Table 9.2: Type-token ratios (%) across the PV groups**

<b>PV group</b>	<b>CLEC</b>	<b>LOCNESS</b>
<b>UP</b>	6.2	22
<b>OUT</b>	8.9	24.9
<b>ON</b>	9.5	19.7
<b>ABOUT</b>	20.9	20.5
<b>DOWN</b>	14.8	37.5

Third, there is a partial overlap in the high frequency PVs in CLEC and LOCNESS. The five most frequent PV types across the five groups are presented in

Table 9.3, with the types shared by both corpora underlined. Those with frequencies below ten times are not included in the table due to the low numbers. Compared with the most frequent types of the German and Italian learner corpora (Table 9.4, adapted from Waibel, 2007: 87), a number of PVs are common across all of the corpora, including: GIVE UP, MAKE UP, CARRY OUT, GO OUT, FIND OUT, GO ON, TAKE ON, and BRING ABOUT. These eight PVs are of great importance, since they are prevalent in both the native and non-native languages, and six of them have been analysed in this study. Among them, the register used with CARRY OUT is deemed more academic, while those with GO ON and GO OUT are more like speech (Waibel, 2007: 95). It has been stated that the use of both formal and informal PVs exist concurrently in LOCNESS and CLEC (Waibel, 2007: 116). This is further supported by Liu's (2012: 31) finding that although PVs are not generally commonly used in academic writing, a selective group of them (e.g. *make up*, *carry out*, *go on*) are frequently presented. As a result, describing learner language as more colloquial (see Chapter 3) by drawing on evidence of the presence of informal PVs appears to be inappropriate, as native English also contains substantial numbers of informal ones. We need to confirm that the use in a learner corpus is far more over-represented, and detailed qualitative analysis must be done before such conclusions can be made.

Waibel (2007: 95) explains the reasons for the abundance and difference of the most frequent PV types in learner corpora as being that topic effects or the influence of L1 and teaching are at work, which is in line with the results of this study.

**Table 9.3: The five most frequent PV types across groups in LOCNESS and CLEC**

UP		OUT		ON		ABOUT		DOWN	
LOC	CLEC	LOC	CLEC	LOC	CLEC	LOC	CLEC	LOC	CLEC
bring	get	<u>carry</u>	<u>go</u>	<u>go</u>	<u>go</u>	<u>bring</u>	<u>bring</u>	break	knock
end	<u>give</u>	point	<u>find</u>	<u>take</u>	get/keep				sit
<u>give</u>	use	<u>find</u>	<u>carry</u>	carry	<u>take</u>				fall
grow	<u>make</u>	<u>go</u>	put		put				write
<u>make</u>	take	get	jump/take		live				look

**Table 9.4: The most frequent PV types of G-ICLE and I-ICLE (Waibel, 2007:87)**

UP		OUT		ON		ABOUT		DOWN	
G-ICLE	I-ICLE	G-ICLE	I-ICLE	G-ICLE	I-ICLE	G-ICLE	I-ICLE	G-ICLE	I-ICLE
<u>give</u>	grow	<u>find</u>	point	<u>go</u>	<u>go</u>	–	<u>bring</u>	sit	fall
get	bring	turn	<u>carry</u>	put	keep				
wake	<u>give</u>	<u>go</u>	<u>find</u>		carry				
bring	<u>make</u>	point	turn						
end	build	<u>carry</u>	<u>go</u>						

Finally, a minor finding is that some cases show variations in word form distribution, such as CARRY OUT and the comparison of FIND and FIND OUT in Chapter 7. The preference for different word forms by different groups of speakers bears out the fact that a particular word often has a predominant meaning matched by a particular word form (Sinclair, 1991), and the Chinese learners diverge from NSs in their penchant for particular word forms. The results also bring up the risk of considering quantitative data alone without detailed analysis, and demonstrate that different verbs may display a tendency towards one or more particular word form.

To recap, the overall results of the distributional differences reveal that:

1. The Chinese learner corpus CLEC has more occurrences of PVs due to the repetition of using the same types.
2. The NSs have more diversity of PV types, as shown by the type-token ratios.
3. Some of the most frequent types of PVs in the two corpora are shared, including formal and informal PVs. The most prominent PVs in CLEC are, like other learner corpora, affected by topics/genres/registers, mother language and instruction.

4. Preference for word forms can work to characterise the behaviours of a particular phrasal verb, but they must be probed in depth to ensure accuracy.

### **9.2.2 Question 2: Idiomaticity and restriction strength of PVs**

When it comes to the study of collocates, two specific features, degree of idiomaticity (opaqueness) and restriction strength, have been noted to play crucial roles in categorising collocations (cf. Chapter 6). As concluded from the literature (see Chapter 2), most research on PVs addresses the importance of idiomaticity and collocation restriction, however, they are taken as two separate concepts in some studies and sometimes combined as one notion in others. It has been argued that these two issues do not necessarily suggest each other (see Section 6.4). More degrees in one do not imply more degrees in the other, although a tendency is often observed. Since these two issues are independent, I proposed to produce an inventory in which we can locate PVs by their restrictedness and idiomaticity. A two-axes graph was drawn (see Section 6.4.2) to account for the properties of PVs, at the same time highlighting that there is not necessarily a match between idiomaticity and restriction of collocations.

This graph foregrounded the fact that learners' difficulties with PVs are two-dimensional (idiomaticity and collocability) and these issues must be dealt with

separately. This is a question relating to learners' task being production (productive) or comprehension (receptive). As seen in Chapter 2, learners' major difficulty is usually ascribed to the idiomaticity of PVs. Research studying PVs in learner language often concludes that idiomaticity is a major factor which brings problems to language learners, but many of these studies are only testing the comprehension ability (receptive knowledge) of learners, for example the experimental methods applied by many researchers (Dagut & Laufer, 1985; Liao & Fukuya, 2004; Yoshitomi, 2006). A few researchers who study learners' L2 writing also claim that the idiomaticity of PVs results in learners' low achievement. For example, Lam (2003: 177) finds that Cantonese learners make more errors in what he terms 'semi-figurative VPCs', which are parallel to the 'semi-transparent PVs' in this present study. The most figurative ones (= idiomatic PVs here) are also found difficult, but not as much as the semi-figurative ones. However, his findings are based on errors comprising various kinds of misuse, and most of them are actually mis-collocations from the view of this present study. As such, learners' misuses of PVs are not a consequence merely of idiomaticity; collocability must also play a role.

Indeed the idiomaticity of PVs causes problems for decoding, and learners are easily blocked when they come across a semantically opaque PV. However, for

encoding, i.e. producing, text, when they are writing, they have options to choose a more transparent, general, basic substitution, paraphrase the PV or simply avoid using it. As a result, if the production of text, i.e. writing, is to be improved, attention may be better directed to the collocability of PVs first, as informed by this two-axes graph showing that greater idiomaticity does not inevitably give problems to learners, and it is the uncertainty of selecting collocations which makes learners fall victim to misusing PVs.

Since the potential problem of idiomaticity can be forestalled as long as the meaning is made clear to the learner, it is the problem of collocation restriction that will be the major source of challenges to the learners. A transparent phrasal verb with free collocation may not be problematic to learners either, because the meaning can be easily guessed and the collocation is wide-ranging. The most problematic items may be the categories in between, as pointed out by Nesselhauf (2005), who states that neither the free collocations nor the fixed collocations, but instead the restricted collocations, will cause problems for learners. Consider the example DRAW UP + [ideas] in the opaque-and-restricted group of the two-axes graph. Not all words referring to written forms of ideas can be collocable with it, for example: *?draw up a fiction*, *?draw up a blog*. Non-natives will find the collocations less predictable than

they expect, even when the semantic field [ideas] is delineated. Such semantic fields must be refined with more details. This finding is in accordance with Cowie and Howarth's (1996:86-87) work, in which their subject of a novice learner is found "grappling with broad semantic distinctions rather than managing collocational niceties", e.g. \**expanding abilities* instead of *extending/developing abilities*. The requirement of the object of *expand* is 'a volume or space'; the noun *abilities* does not belong to this semantic field, thus is disqualified. Such a lack of precise prediction of semantic requirements is a great hindrance for learners.

Idiomaticity alone is not an insurmountable problem for learners. As a matter of fact, an idiomatic PV with fixed collocates does not puzzle learners much, as long as it is taught. Although learners do not encounter considerable difficulties in using idiomatic PVs, they suffer from PVs which have limited collocation candidates confined by a semantic field.

Coincidentally, a similar idea of allocating multi-word expressions to a framework based on transparency/opaque and frequency has been proposed by Martinez (2013). The two criteria were devised to prioritise multi-word expressions to determine their order of teaching. Martinez's proposal cannot escape the problem of subjectivity either, but he argues this could be minimised by setting thresholds of

frequency with corpus data and the learners themselves judging the transparency. Martinez's framework and the model of this present study together reflect that distributing multi-word units (including PVs) on continua of their features can be envisaged to be useful to teaching/learning. Follow the same line of thought, instructors will be empowered to create more effective teaching plans.

### **9.2.3 Question 3: Discriminating a PV and its synonym**

During the analysis of PVs in the Chinese learner language, the problem of confusion with its single-verb equivalent emerged. In some instances of PVs, the addition of a particle does not change the whole meaning drastically, but the similarity of forms certainly puzzles foreign language learners. However, researchers such as Partington (1998), among many others (see Chapter 7), have successfully demonstrated that even synonymous words can be characterised by discriminating phraseological patterns. Accordingly, a verb and a similar phrasal verb can also be distinguished by their phraseologies, which will help learners clarify their differentiations. Discriminating a PV and its synonymous counterpart is believed to help foreign learners tell them apart.

The examples FIND and FIND OUT were selected for this purpose. Their phraseological patterns in BoE were investigated first to demonstrate the

distinguishing characteristics of their usages (see Section 7.4.3). A picture of these two verbs' patterns reveals the fact that they share many patterns but also differ in a few points. I moved on to analyse these distinct patterns in the Chinese learner language but obtained no erroneous uses. An alternative approach, seeking the Chinese learners' errors in the concordance samples, was adopted. This successfully brought to light that the Chinese learners misused the phrasal verb FIND OUT in places where FIND is the accurate choice. It has been suggested that the reason for this is that the Chinese learners have poor knowledge of the linkage between the PV and its possible semantic fields. Such findings lend further support to the importance of introducing more extended lexical units or phraseological units to learners, as these will aid in discriminating items which share semantic and formal similarity.

#### **9.2.4 Question 4: The Chinese learners' uses of PVs, with a focus on phraseologies**

A number of gaps have been noticed in the literature on PVs, phraseologies and learner language (Chapters 2-4). Although the significance of the phraseological units to a word has been well established, its application to the description of learner languages has been attended to mainly in cases such as idioms, formulae and VN combinations, etc., instead of the phraseological behaviours of a lexical item. In

particular relation to PVs, mainstream previous research has tackled the senses of particles and the classifications of PVs. Work examining learner languages also places more emphasis on the possible combinations of the verb and the particle rather than the phraseological environment and the wider context where a PV is used. Yet, to successfully learn an item, say a PV, means to internalise the usage constituted by all of the ‘extended lexical units’ in respect of the PV in question, thus this thesis has focused on giving accounts of the patterning behaviours of PVs in the Chinese learner language, as outlined below:

#### *9.2.4.1 Semantic and syntactic preferences: collocations and colligations*

Previous lexical studies of learner languages have generally indicated that learners produce many mis-collocations (Howarth, 1998; Kaszubski, 2000:33; Nesselhauf, 2005). While this type of research has gained many insightful results in the area of errors, this present work attempted to discover the Chinese learners’ idiosyncratic selection of collocates. In other words, I surveyed the areas where collocations made by the Chinese learners are different from those of native speakers.

The results of the five PVs in Chapter 6 brought into view that NNSs every so often put together a few peculiar nouns with these PVs but NSs are able to draw on

more varieties of words, which may be regarded as ‘difficult’ collocates to learn at advanced level from the viewpoint of learners. This suggests that the Chinese learners need to boost their knowledge of more collocable nouns, especially those which are arbitrarily constrained. The divergences of two-word collocation have also been evidenced by De Cock’s example of *good fun* (2011:208). The two words are frequently collocated in English but no such collocations are found in all the learner corpora (German, French and Chinese). Evidences from other PVs (e.g. CARRY OUT and TAKE ON) also showed that, unlike native writers who can bundle with diverse words, the Chinese learners appear to be restricted to a limited range of collocates. The limitation can be witnessed in the nouns which fit into a slot in a string. ‘*It is [evaluation] to find out*’ was found most frequently in English, with a number of words filling in the [evaluation] blank; however, limited variety was shown in the Chinese LL, which only contains one fixed word, *possible*. Such a preference may be explained by the inclination towards safe pieces, as noted by Kaszubski (2000:241) and Nesselhauf (2005:69). Or more likely, when the Chinese learners intend to use FIND OUT, *it is possible* is primed, as Hoey (2005) suggests, and becomes the first choice for the Chinese learners to use.

Besides the collocations of PVs and nouns, there is also a prominent learner

preference of one structure over another. In the example of FIND OUT, the Chinese learners adhere to the VN pattern while natives incline towards Wh-words (Chapter 7). Also it was found that the pattern 'FIND OUT + *about*' is not presented in the Chinese learner language, while this pattern is attested as a non-peripheral one in general English. These lexical collocations and structural colligations imply that Chinese learner language is idiosyncratic and not rich in expressivity.

#### *9.2.4.2 Semantic sequence and prosody*

Besides the two-word units, larger strings of text were also considered. In fact, these are the paramount concern of this thesis. These longer strings have been identified in the shape of semantic sequences made of fixed and flexible components. At this point, I summarised the significant results, by the method of describing the characteristics of the Chinese LL.

**1. The Chinese learner language lacks certain typical English expressions while at the same time displaying certain learner-exclusive patterns. Moreover, the linkages of some prominent semantic concepts with a PV in English are not associated by the Chinese learners, and the Chinese learners often connect non-native-like semantic/conceptual elements to a particular phrasal verb.** Some typical sequences in native

English are found absent or occurring in relatively low frequency in the Chinese learner language. Certain sequences are prevalent in native data while they do not occur at all in the Chinese learner data (as seen in Section 7.3 CARRY OUT, Section 8.3.2 COME ABOUT, Section 8.4.2 CUT DOWN), so is the case of BRING UP, of which one semantic sequence is found only in native writings. Furthermore, there are some examples where LOCNESS do not show any patterns but CLEC manifest learner-specific ones (see Section 8.3.1 BRING ABOUT and Section 8.4.2 CUT DOWN).

In addition, there seem to be ‘conceptual differences’ which can be observed by the uses. Examples are GO ON (Section 8.2.1), BREAK DOWN (8.4.1), GROW UP (6.3.1.4), BRING UP (6.3.1.3), CARRY OUT (7.3.1.3), TAKE ON (8.2.2). What is found emphasised differently by the two groups of writers is summarised in Table 9.5. In these cases, what one phrasal verb is associated with subsumes various units: some look like semantic fields, some are like semantic prosody, and some are hard to define thus are taken as different concepts. We will discuss the possible reasons of these differences in further details in Section 9.3.

**Table 9.5: PVs' different emphases by the Chinese learners and native students**

	<b>CLEC</b>	<b>LOCNESS</b>
<b>GO ON</b>	amounts or intensification	--
<b>GROW UP</b>	results	progress
<b>BRING UP</b>	process	purposes
<b>CARRY OUT</b>	public affairs	action/activity
<b>TAKE ON</b>	difficulty	quality

2. **The Chinese learner language also exhibits sharp contrasts of semantic prosody in some cases.** This is demonstrated by instances such as TAKE ON (8.2.2), which usually implies negative connotation, as well as BRING ABOUT (8.3.1), to which NNSs associate unfavourable prosody, expressed by nouns like *problem/mistake/failure*.

3. **When expressing similar meanings, the Chinese learner preference in each example is noted. Sometimes PVs are used by the Chinese learners to replace similar single-word verbs and vice versa.** It has been suspected that the Chinese learners are enticed to SET UP + *constitution* while the NSs use DRAW UP + *constitution*. In sentences where the meaning 'to compel observance of the law' is intended, the Chinese learners prefer CARRY OUT instead of ENFORCE, which is more often selected by the natives. Besides the

substitution of single-word and two-word verbs, the Chinese learners are also more likely to select the word *continuously* rather than the native preference ‘GO ON *for time*’. It is also conjectured that the NNSs use PUT/RAISE+ *question* in similar senses to BRING UP + *issue/point/question*. This thesis took an alternative approach in looking into each instance, and verified that Chinese learners make use of other verbs in contexts where a phrasal verb is likely to reside.

**4. In some cases, the Chinese learners appear to adhere to literal rather than idiomatic uses of PVs.** For PVs which have both literal and idiomatic senses, it was often found in the Chinese learner data that the literal uses overwhelmed the idiomatic ones, as evidenced by BRING UP and PICK UP (Chapter 6). The literal uses were also noted to be privileged in the Chinese LL as demonstrated by CUT DOWN (Chapter 8).

### **9.3 Discussion**

In Section 9.2 and the results chapters, we have seen that the Chinese learner language has characteristics which include limited range of collocations, absence and preference of structures, absence and preference of phraseologies and semantic prosody, non-native-like uses of PVs and single-word verbs, and a favouring of literal

uses. The following sub-sections will provide a general discussion of the possible sources of these findings, and of the meanings of the findings and how they can add to our knowledge of phrasal verbs and learner language.

### **9.3.1 Influence of topics, genres or registers**

From the findings summarised above, factors such as topic, or genre/register appear to have a significant influence on PVs usage, and thus deserve some discussion at this point. Of the phraseologies which are found exclusively in one corpus, some seem to be influenced by differences either in genre/register or in topic between the two corpora. Although no genre or register analysis is conducted in this study, the influence of genre/register can still be noted in some instances. For example, the wide uses of BRING UP + [subjects] in LOCNESS but not in CLEC appear to be a result of genre/register disparity, as LOCNESS is largely composed of argumentative essays whereas CLEC is composed of examination essays, diaries, letters and so on. The CLEC-only instances such as BREAK DOWN + [goods of poor quality] are clearly evidences of topic influence (the topic 'My View on Fake Commodities' is selected in CLEC; see Chapter 5). In some cases, then, topics and genres/registers of the essays certainly influence which PVs will be used most frequently. Specific topics or genres/registers will not only prompt specific PVs but also select for literal or

figurative usages. Literal PVs such as GET UP, RISE UP (Section 6.2) and PICK UP (Section 6.3) are more likely in some genres (e.g. descriptive writings including anecdotes, and diaries); figurative PVs such as BRING UP, and END UP are often used in formal texts such as expository and argumentative essays (see Section 6.2).

The preference for literal PVs in CLEC found in this thesis (see point 4, Section 9.2.4.2 above), is probably a consequence of genre, as CLEC contains some texts of diaries and letters, in which literal uses of PVs are likely to be frequent. These observations are consistent with those in other studies, such as Hinkel's (2009) discussion on topic effect on features of L2 texts; see also Biber et al. (1999) and Liu (2011), where certain PVs (e.g. CARRY OUT, BRING ABOUT) are found to be more frequent in academic writing. This influence of topic and of genre/register on the research findings means that conclusion about comparable native and learner corpora have to be treated with caution. The absence of a particular phrasal verb in the learner corpus, for example, may be a result of the tasks the learners were asked to complete rather than a lack of knowledge on their part. However, as no truly comparable native/learner corpus yet exists, we simply have to bear these differences in mind when drawing conclusions from the research.

### **9.3.2 L1 influence**

Another crucial factor is L1 influence, and I will discuss two major aspects of this factor. The first aspect relates to the structure differences between L1 and L2, and the second aspect relates to the phraseological differences revealed by this present study. Thus, the discussion presupposes a distinction between the structure of a language (presence or absence of PVs) on the one hand and its phraseology (collocation, semantic prosody and so on) on the other.

The structure differences between L1 and L2 have already been mentioned in Section 3.4.2, where it was noted that some researchers suggest the absence of PVs in L1 will cause difficulties for learners to use PVs in L2. In other words, it is necessary to ask whether a first language which does not have PVs will create difficulties if a speaker of that language learns a second language that does have PVs. In previous studies, for example, the avoidance of PVs by learners of English with Hebrew and Finnish as L1s has been reported to be a consequence of the lack of PVs in these two languages (Dagut & Laufer, 1985; Sjöholm, 1995). However, our results of PV frequencies presented in 9.2.1 show that the Chinese learners in the study reported in this thesis do not refrain from using PVs. As Chinese is also a language which lacks the structures of PVs, this present study does not provide support to the claim that the

structural influence of learners' first language will necessarily result in learners' difficulties. We need more evidence from other L1s which also do not have PVs in the future, before any convincing conclusions of L1 structural influence can be drawn. So far, it is only safe to say, at least for Chinese learners, the absence of PVs in their L1 (Chinese) appears not to affect their L2 (English) significantly.

With respect to language structure differences, there is one point which is also worth mentioning: the use of PVs and single-word verbs. Previous research such as Waibel's (2007:88-89) which has touched on the paraphrases of PVs and single-word verbs by comparing the frequencies of GO ON (vs. CONTINUE) and BRING ABOUT (vs. CAUSE) finds learners' preference for PVs. The much higher frequencies of PVs suggest that German and Italian learners consistently rely more on PVs, in contrast to British/American students' preference for single-word verbs. However, the findings in this thesis do not suggest a consistent pattern of preference for either PVs or single-word verbs among the Chinese learners. Instead, my study suggests that each PV must be considered separately, on a case-by-case basis (see Section 9.2.4). In light of my findings, the influence of the L1 structure is not clear with respect to the uses of PVs or single-word verbs, since the Chinese learners in this present study do not have a tendency towards using either of the verb types overall.

The second aspect is the fruitful discovery of non-native-like phraseologies in this thesis. The influence of L1 has been seen to be conspicuous in the phraseologies analysed from Chapter 6 to 8. We can see not only that word-for-word direct translation of L1 can transfer to L2, but also that associations of concepts in the Chinese learners' English are built largely on Chinese concepts.

The unusual collocations found in CLEC, for example KNOCK DOWN + *ground*, indicate that some collocations are liable to be affected by the first language, Chinese (see Section 8.4). The findings also suggest that the degrees of idiomaticity of the L1 translation may contribute to the unconventional use in their L2 English, as evidenced by CARRY OUT + *law* (see Section 7.3.1.2). This endorses the previous studies of collocations, where the congruence or non-congruence of L1 and L2 is confirmed to play a crucial factor (Nesselhauf, 2005:221-229). The results of the synonym comparison, i.e. FIND OUT (Section 7.4.3), also demonstrate that the L1 may hamper the Chinese learners' accurate use of the two synonymous verbs because their first language, Chinese, has only one equivalent verb, 'f ā xiàn', which can represent either the PV or the single-word verb. The anomalous collocations have already been found in studies such as those seen in Section 9.2.4.1. More specifically, inappropriate collocations of PVs have also been discerned in the LLs of Italian and

German L1 speakers (Waibel, 2007).

More surprisingly, this present study reveals that the impact of first language not only is found in collocations, but can also be discovered in other phraseological phenomena (see Section 9.2.4.2). This is significant for our understanding of learner language, as phraseological units other than collocations have rarely been explored before. Many cases in our findings show that the Chinese learners relate the PVs to some idiosyncratic concepts or connotations, which probably can be traced back to the first language, Chinese. The L1 transfer seems to be at work not only at concrete (word-word congruence) but also at abstract (concept congruence) levels. For units larger than words, not many studies can be compared because this field is still waiting to be explored further. One of these studies is Paquot (2008), where the author reports that more variable phraseological units are influenced by the French learners' L1. Unfortunately, this study looks at only continuous sequences (e.g. *take the example of*) but not any more abstract elements like those found in the present work. What has been found in this thesis can be added to studies of learner phraseology. Learners' L1 can manifest itself at all levels of phraseological units, including both word-word concatenations and word-concept associations.

The results of the present study have also successfully identified the semantic

prosody differences between the Chinese LL and English (see Section 9.2.4.2). My study is consistent with that of Xiao and McEnery (2006:125). In Xiao and McEnery's contrastive analysis of semantic prosody in Chinese and English, they obtain mixed results in that while many cases of near-synonyms in the two languages they studied have similar semantic prosodies, some show differences. They explain the similarity with the commonality of the human conceptual system suggested by Sweetser (1990). They ascribe the differences to the dissimilarity of the lexicons (which reflect language-specific concept structures) in the two language varieties. Learners are indeed less sensitive to the prosody existing in English. McGee (2012) tests the semantic prosody awareness of NSs and NNSs, and proves that for many cases, NNSs have not been able to pick up the implicit prosody shared by NSs. Reflecting on these previous research and my study, we can thus surmise that the Chinese learners' perception of semantic prosody in relation to a lexical item is, at least partially, not like that of natives.

As to the final point of question four, our results also point to the same noteworthy preference for literal PVs over idiomatic ones in the Chinese LL, as has also been attested to in Finnish and Swedish learner languages (see Section 3.4.2). The favouring of PVs/single-word verbs and literal PVs may have little to do with L1

transfer and have more to do with the influence of genre/register types as discussed earlier. It may also be a consequence of learners relying on items which they know to be 'safe' or familiar. Our findings of the Chinese learners' preferred use of these basic or simpler PVs supports Yorio's (1989:64) report that advanced L2 learners employ two-word verbs in a similar proportion to native speakers, but learners tend to use those which are less idiomatic and grammatically simple.

My study is, of course, restricted to Chinese learners of English, and the results cannot be generalised beyond this study. The results do not apply to all learners of English, though they may be considered reasonably accurate for all Chinese learners of English. At the same time, it is worth considering to what extent the study and its findings might be deemed relevant to speakers of other languages. The first point to make is that the methodology and the questions of the research in this thesis might certainly be transferred to learner corpora produced by learners with other L1s.

Secondly, it is reasonable to predict that the differences in phraseology and in semantic prosody, similar in kind if not in detail to those found in this research, will also be identified in all learners, not just Chinese ones. These idiosyncrasies for the Chinese learners seem not to be language-specific; that is, they do not exist because the learners' first language is Chinese. Any other languages surely have combinations

or associations not consistent with the conventions in English. However, the specific points which have been found in the instances analysed in this study are unlikely to be generalisable. We have also seen above that other studies have provided some evidences of similar phenomena from other L1 backgrounds. Investigating the phraseologies of PVs in other L1 groups will very probably also identify collocates and phrases which are influenced by the L1, as those of the present study. Finally, though, we must note that the presence or absence of PVs in an L1 alone is not an accurate predictor of the frequency or accuracy with which learners produce PVs in English. In the next section, the theories of L2 lexicon and conceptualisation will be used to account for these learner idiosyncrasies.

### **9.3.3 L2 lexicon**

Taken together, the phenomena described in points 1 to 2 (Section 9.2.4) above may be accounted for by Kroll's (1993) representations of learners' lexicons and Danesi's (1995) 'conceptual fluency', both cited in Lam (2003:50, 53-54) to explain L2 lexicon. Kroll produces two models of beginners' and experts' lexicons, and proposes that learners are moving from a primitive state where the lexis of L2 is related to concepts through L1 and, towards a better state where L2 directly links with concepts. Earlier

accounts of L2 vocabulary acquisition have supported this view. Wolter (2006:744) argues that the first language lexicon provides L2 learners with a pre-set structure of concepts; the dissimilarities of the first and second language lexicons may result in miscollocations in their learner language. He explains that learners make collocational errors because they rely on their L1 lexical knowledge, but acquiring the new combination of words in L2 will lead to ‘conceptual modification’, posing problems to the learners. This idea is endorsed by Danesi’s conceptual fluency, which argues that “students ‘speak’ with the formal structures of the target language, but they ‘think’ in terms of their native conceptual system” (Danesi, 1995:5). Although Danesi focuses largely on metaphorical ideas in languages, the application of conceptual fluency to the findings of this thesis is likely to be reasonable.

Besides the first language interference, another rational explanation may be that learners may lack the same consensus of phraseologies that every native speaker holds subconsciously. As pointed out by Wray (2002:206), the L2 learners’ problem is that they have ‘too much choice’, so have no idea of how to choose the appropriate word from a number of grammatically possible words. Irrespective of how the mechanisms for selecting and combining words operate in their minds, we can be confident that the learners’ lexicon is dissimilar to the native lexicon. The different mental lexicon

structures between L1-L2 languages bring about the unusual characteristics of learner language, that is, its un-naturalness.

The models of L2 lexicons in the literature illustrate the ways in which languages are organised and processed in our mind and consider only individual lemmas (for details of the models, see Singleton (1999:84-110)). Recently some researchers have used the term 'phrasicon' to refer to the phraseologies of language (Paquot & Granger, 2012), but they have not addressed the storing and structure of the phrasicon. The L2 mental lexicon (including both vocabulary and grammar) is constructed by interrelated linguistic items which make a massive intertwined network. The associations between linguistic items are activated and strengthened each time learners come across the usage pattern, and finally they learn or acquire the usage (R. Hudson, 2008). L2 learners at a less proficient level may 'prime' (in Hoey's term) items which are not coming up readily in the native writer's mind in the same situation. Whether this is a consequence of L1 interference or an instruction effect, the learners become 'fossilised' (Bybee, 2008): when their minds search for the extensive units of a word's usage, the non-native combinations are recalled first.

Nesselhauf (2005:288) has argued that a model based on the strength of links can more adequately explain the learner's lexicon than one which sees the lexicon as

comprising words and chunks. She concludes that “the links between elements of (semi-)prefabricated units are weaker in the advanced learner’s mental lexicon than in the lexicon of native speakers--both the links between the elements of collocations and those between collocations and larger units of usage” (Nesselhauf, 2005:288). The learners’ phraseologies obtained in this study are indexical of the L2 lexicon, and substantiate Nesselhauf’s claim by providing more evidence of small and large patterns.

It is widely accepted that the development of the L2 lexicon is a continuous process of refining the meanings of individual words (Sonaiya, 1991:274), and “lexical units are increasingly processed qua meaning than qua form”, as concluded by Singleton (1999:189). Learners have to shape their lexicon not only by disambiguating the meanings over and over again, but also by rebuilding and reorganising the lexical items in the lexicon. As recorded in Chapters 6 to 8, it has been ascertained that the Chinese learner language is prefabricated to some extent, but that the Chinese learners associate different elements which are rarely or never combined by L1 writers. The network of the L2 lexicon has a structure different from that of the native lexicon, hence the learners’ task will be to reset or strengthen/weaken the links between the lexical items to match the native target. In

fact, not only lexical items but also the related concepts must be fine-tuned bit by bit, as the phraseological units of a lexical item clustered by learners and natives may not match.

To summarise, this thesis has dealt with four aspects of PVs in the Chinese learner English and native English: their frequency distribution, their properties of idiomaticity and collocability, the issue of synonyms, and, most importantly, their phraseological units. Earlier accounts of formulaic expressions or prefabricated units have proposed the tentative statement that L2 learners use these chunks no less than native speakers. In other words, learners are capable of applying the idiom principle (Sinclair, 1991) to an extent comparable to NSs (Weinert, 1995; De Cock et al., 1998; Lesniewska, 2006:101). The Chinese learners in this study confirmed that their learner language is idiomatic, as a great extent of consistency and systematicity has been noted.

Despite the Chinese LL being idiomatic, the phraseologies of the Chinese LL have been shown to be largely distinct from the native writers'. Case studies of PVs were conducted to examine the individual PV's collocational behaviours, from single units (their collocates) to the 'extended units of meaning' (Sinclair, 1996), such as multi-word units (patterns, semantic sequences) and more abstract notions (semantic

preferences or associations). Starting with collocates, the selected examples of PVs were found to produce different semantic sets from the Chinese learners as compared to natives. At a higher level, the semantic sequences of the PVs were identified to see whether the typical phraseologies in relation to a PV in the two corpora are the same. Interestingly, many of the PV examples studied in this thesis show distinct semantic sequences across the corpora, suggesting that the Chinese LL bears little resemblance to standard English in the collocation of a PV and its habitual words. These phraseological units can therefore be adopted as a means to reveal learner-specific features. An interpretation of the divergence is that perhaps the networks of the Chinese learner language and English are differently structured and the inter-connection strength between lexical items varies across languages. For learner patterns contrasting with those in English, the Chinese learners seem to have their own distinct network of phraseological units. The reason for these learner-specific phraseologies may be that their knowledge of the most relevant associations to a word is not like that of the native students'. Above all, this thesis has confirmed that learner language idiosyncrasy can be discovered in respect of phraseology. The contextual approach, which applies the notions of phraseological units to the learner language study, can bring to light what has been overlooked before.

In addition to those reasons discussed above, we can mention in passing other plausible factors that may also be responsible for the missing or skewed phraseologies. For example, some of those phraseologies which are used by the NSs exclusively appear to be 'higher-ordered' materials which the Chinese learners may have not encountered before or had the chance to use, as they are at lower levels compared to native writers.

Since the phrasal verb features in CLEC and LOCNESS have been identified, the results can inform the teaching and learning of English phrasal verbs. A series of pedagogical recommendations thus can be advised accordingly as below.

#### **9.4 Pedagogical recommendations**

In this section, I will put forward six pedagogical recommendations based on the findings of this thesis. These recommendations are presented in the order as below, and will be discussed individually in six subsections:

- Classing PVs is not absolutely necessary
- PVs with an identical particle can better be presented together, in texts of specific topics, notions, genres or registers
- Selecting appropriate PVs for teaching
- The two-dimensional model of PVs and its implications for teaching

order

- Contrasting PVs in conjunction with the synonymous one-word verb
- Focusing on the phraseologies of PVs

#### **9.4.1 Classing PVs is not absolutely necessary**

A significant issue of teaching PVs is whether it is necessary to separate PVs from other vocabulary items as an independent and special group. To make a better decision, we have to consider two questions: are PVs different from other vocabulary items and do PVs need different treatment in syllabuses and textbooks? PVs stand out from other words due to the regular verb + particle forms, which are noticeable to both teachers and students, and thus may be regarded as a unique constellation. Also, the instructors may like to class PVs apart from other vocabularies in order to emphasise PVs and familiarise students with them. From these perspectives, PVs are inevitably distinguished from other vocabulary items.

Nevertheless, the present study shows us not only that the phraseological behaviours of each phrasal verb are idiosyncratic in contrast to native English (9.2.4), but also that each PV is individualised, rendering assembling all PVs together less sensible. Given the specific behaviours of individual PVs, presenting all PVs to learners seems to be not very descriptively useful. Isolating PVs also leaves learners

with a false impression that PVs are special phenomena, which are unlike common items of vocabulary. The fact is that learning PVs requires gaining knowledge of the phraseological information from the context, and so does learning common words. In particular, learning PVs especially requires such information because many PVs appear to have special usages. For example, CUT DOWN (+ON) has a negative prosody, but the synonymous common verb REDUCE does not (see Section 8.4.2). These figurative meanings or connotations have to be made clear to students. As indicated, phraseological information is crucial for learning both common words and PVs. Since the teaching method of including context could work for PVs and other words, there seems to be no compelling reason to separate PVs from the vocabulary list, and thus I do not strongly advocate classing PVs separately.

#### **9.4.2 PVs with an identical particle can better be presented together, in texts of specific topics, notions, genres or registers**

The current trends of phrasal verb teaching and learning have been reviewed in Sections 3.4.2 and 3.4.3. Basically, the coverage of grammatical rules, semantic features, the combination of the verb and the particle, pragmatic appropriateness, and so on, have been emphasised by different researchers. As to the instruction methods for phrasal verbs, they are largely informed by cognitive linguistics. Mounting work

(Dirven, 2001; Kurtyka, 2001; Yasuda, 2010; Pozdnyakova & Gunina, 2011; White, 2012) has advocated drawing on cognitive strategies to assist students in learning phrasal verbs. Metaphoric schemas (e.g. spatial orientation) or semantic/syntactic groupings are advisable means in the teaching procedure. In practice, teachers could show students the basic and derived meanings which can be illustrated by visual images (Kurtyka, 2001:46). Besides the conceptual approaches stated above, some scholars would like to suggest that instructors guide students to notice that certain PVs are related to certain topics (e.g. *boil over* and *chop up* are related to 'food') (Pozdnyakova & Gunina, 2011:357), which also seems effective for teaching and learning. I concur with these researchers that cognitive approaches can aid phrasal verb acquisition by aligning conceptually similar items together.

Undoubtedly, PVs should be taught with methods which can facilitate the learning process. For example, PVs can be presented according to themes, particles or verb meanings, as suggested in Lam (2003: 218). The present thesis provides insights particularly for grouping PVs by particles and relating PVs to themes/notions/topics/genres/register. We have seen the advice given by Quirk et al. (1985) and Sinclair (1991) in Section 2.5.1 that grouping PVs by their particles is useful because parts of their meanings can be explained by the analogous figurative

meanings of that particle. The present study also adopts the same approach by analysing PVs with the same particle together (Chapter 6 to 8 deal with different particles). Also, it supports the notion that that this approach is beneficial if a number of PVs with the same particle are simultaneously accounted for based on their features, as seen in Figure 6.1.

In addition, the present findings also provide insights into relating specific PVs to specific notions or genres/registers. Some PVs seem to be more probably linked to certain text topics or notions; for example, DRAW UP can be introduced to students when discussing the Constitution, and BREAK DOWN in texts involving computers or fake commodities (see Chapters 6 and 8). There are tendencies for different PVs to occur in different genres/registers as well, as we have seen in Section 9.3.1. Pointing out to learners these tendencies for specific PVs to feature in specific notions and genres/registers would be beneficial.

### **9.4.3 Selecting appropriate PVs for teaching**

Next we should consider the parameters for selecting PVs. Determining what PVs should be included in teaching materials is not as easy as it seems. Nesselhauf (2005: 256-260) proposes a three-dimensional approach which rates collocations by their frequency, difficulty and disruption degrees, with the goal of identifying useful

collocations. This model, as Nesselhauf (*ibid.*) warns, is not satisfying in some respects: the degrees of the three dimensions cannot be determined absolutely, and it is better applied to courses for advanced learners only. As this study does not look at parameters other than frequency, we will only discuss frequency here.

One common concern in selecting PVs for teaching is to choose those with great frequencies such as the 100 most frequent PVs identified by Gardner and Davies (2007). Nevertheless, my study suggests that remarkably frequent PVs may not be of great instructional value at all times. There are two reasons for making this pedagogical claim. First, the very frequent PVs often comprise multiple senses; for example, MAKE UP and TAKE UP (rank 4 and 5 in CLEC) can represent several different senses, and each sense should be listed as a separate entry in textbooks. They will surely confuse beginners and are better saved for advanced learners at later stages. Second, the significance of the usage patterns which have been discussed in this study compels teachers to consider such important information in conjunction with individual PVs. The problem is, however, that many of the patterns must be recognized by human eyes, rendering analyses of extremely frequent PVs infeasible. Meanwhile, asking students to familiarise themselves with too many patterns in relation to one phrasal verb certainly will get students into more difficulties and

probably corrode their learning incentives, especially for novice learners. As suggested by this work, it may be better to prioritise PVs with moderate frequencies, such as those examined in this thesis, rather than the very frequent ones.

#### **9.4.4 The two-dimensional model of PVs and its implications for teaching order**

One of the outcomes of this research is the proposed new grouping of PVs using a two dimensional framework (see Figure 6.1), which was created because the previous classifications have not been satisfactory. The detailed summary of the inconclusive classifications introduced in Chapter 2 brings to light the complicated nature of PVs. The intricacy of their syntactic and semantic features causes difficulty in classifying PVs. This present research has provided a new perspective on grouping PVs using the two-dimensional framework, which records the semantic complexity of the PVs themselves and the selection of their collocates. These two notions, idiomaticity and restriction strength, could work together to account for PVs. Their relations are commonly implied but hardly stated explicitly in the literature, and highlighting their interactions with examples of PVs, as this work has done, is a first attempt to provide a more comprehensive account. The visual representation of PVs using these two criteria is geared to grouping phrasal verbs, with the advantage that the complexities

regarding these two factors can be simultaneously crystalised.

Such a two-dimensional model can be applied to setting appropriate order of teaching PVs. This research advances pedagogy development by relating the teaching order to the classification of PVs. Teaching order of PVs should be considered in relation to proficiency levels. Analyses of PVs with these two parameters, idiomaticity and restriction strength, from lesser to greater degrees create nine categories of PVs (see Figure 6.1). It is rational to speculate that more degrees of idiomaticity and restriction strength would mean more degrees of difficulty. Therefore I recommend teaching PVs in the order from beginner to advanced levels, according to their predicted difficulty degrees (see Section 6.4.2.2), as the following six groups show:

1. Transparent and free
2. Semi-transparent and free
3. Semi-transparent and fixed
4. Idiomatic and fixed
5. Semi-transparent and restricted
6. Idiomatic and restricted

If greater difficulty degrees could truly be implied by greater strength of

collocational restriction and greater opaqueness of idiomaticity, then we could roughly determine the difficulty degrees of each PV group in the two-dimensional model. On the one hand, the idiomatic collocations are conjectured to be more difficult than semi-transparent and transparent ones. However, on the other hand, the restricted collocations are harder to predict and thus presenting more difficulties than the fixed and free collocations. Therefore I would suggest that teachers treat PVs with restricted collocates (groups 5 and 6) at a later stage.

The transparent-fixed, transparent-restricted and idiomatic-free groups may be taught when the students are in pursuit of reaching native standard, because there are rare cases of these types. The novice learners should start out with the literal and free PVs, because these are potentially more frequent verbs and easy to encounter in texts or speech, and their meanings are fairly clear to learners. Next are the semi-transparent and free PVs, which are also not difficult to understand and use. Those with fixed collocates, either semi-transparent or idiomatic, are moderately problematic to learners, and probably could be presented as idiosyncratic examples which require some memorisation. Classifying PVs with this bi-axis model can help to draw teachers' attention to the PVs in the middle zone: those which have been observed in this study to be more problematic, instead of those which are extremely

idiomatic/literal or strong/weak in binding. The restricted PVs, whose verb components are somewhat opaque, have been underscored in the present study as requiring more attention from more advanced learners. The key point for teaching this type of PVs is to make explicit to students which are collocable and which are not. Data-driven learning, mentioned in Section 3.3.1, is most helpful in clarifying the permissible collocates to learners. The teacher could guide students to create a table like Table 6.8, and to practise dividing semantically similar collocates into semantic fields; this procedure will also simultaneously bring arbitrary collocates to students' attention.

The two-dimensional model is exploratory and can be extended in two ways. The first and most pressing extension is to include PVs with particles other than UP, such as the popular ones presented in Table 2.3. This model can serve for curriculum developers to determine the priority order as discussed above. Further, it can be applied in classroom as exercises. The EFL/ESL practitioners can show a number of PVs with the same particle to students and ask them to locate each phrasal verb in an appropriate place by judging the degrees on the two axes. The second application of this model is to plot targets other than PVs such as verb-noun collocations. For example, the ubiquitous verb MAKE, analysed in Altenberg and Granger (2001: 177),

can be organised based on a similar model: MAKE + *furniture/hole/law* (in the meaning of ‘produce something’) belong to the transparent-free category, and MAKE + *fortune/living* (‘earn money’) can be labelled as semi-transparent-restricted, while the idiom MAKE *it* (‘succeed in doing something’) can be seen as idiomatic-fixed. Such a model could be used to clarify the problems of polysemy associated with both PVs and delexical verbs, thus assisting teachers in classifying some of the most difficult areas of English phraseology.

#### **9.4.5 Contrasting PVs in conjunction with the synonymous one-word verb**

One of the challenges faced by learners is that sometimes a phrasal verb can be synonymous with a single-word verb. Traditionally the synonymous counterpart is used as the gloss of the phrasal verb as a shortcut for students to get the meaning instantly. It is risky to replace them with each other, as this will give students a misimpression that the PV and the synonymous single word are completely equivalent, and so can be swapped without taking the context into account.

Let us consider the case of FIND OUT and FIND in Chapter 7. These two verbs are extremely confusing because their sense explanations and translations are quite

similar in the dictionaries and the learners' L1. Even though the meanings and usages of these two verbs overlap to a great extent, the analysis of this case has illustrated that at some points their behaviours are largely different. I will suggest contrasting phrasal verbs in conjunction with their synonymous friends, pointing out not only the similarities but also the disparities.

Nesselhauf (2005:264) has made suggestions as to teaching, in which systematicity is deemed indispensable. One of the suggestions with respect to systematic teaching is to contrast similar items. Highlighting the dissimilarities between a phrasal verb and its single-verb counterpart can directly raise students' awareness that the two verbs cannot be interchanged at will. As such, teachers are recommended to use corpora as a tool to show differences between the two verbs. Considering a PV and its synonymous partner along with their immediate co-texts is a good strategy for learners to dodge the risks of misuses.

#### **9.4.6 Focusing on the phraseologies of PVs**

Very few of the studies which concern the teaching and learning of PVs have stressed the phraseologies of PVs like the present study. This study argues in favour of taking all available phraseological units into consideration, because these units play an

important part in forming patterns of usage. The various phraseological units identified for each PV in this study demonstrate that they all contribute to the ‘usages’ of one PV, so they must be pulled together to best profile the PV in question. This is in line with Nesselhauf’s suggestion for teaching that “focus[ing] on the exact meaning and usage” is more advisable than “focus[ing] on the form” (Nesselhauf, 2005:269).

The findings of my study not only support Nesselhauf’s statement, but also further reveal the importance of those elements which also shape the meaning but do not have fixed forms. These elements include semantic fields, prosody, semantic sequences and functions, which cannot be learnt by simply memorising possible vocabularies. Students need to be able to clearly know what is permissible in a specific context.

Drawing attention to the phraseologies of PVs, as argued by this work, will lead to the evolution of teaching methods. Traditionally, students are often taught by practising filling in correct answers in context and grouping them on the basis of the particles. Such methods can improve learners’ knowledge concerning one PV by itself, but may be less profitable in terms of the production of texts with a PV. Given the rich results of phraseology difference from this study, revealing the phraseological

behaviours of PVs to students will certainly help overcome the limitations of conventional teaching methods. Furthermore, the individualised patterns of each PV also suggest the need to present their individual properties. This echoes Sinclair's (1991:78) inspiring comment that "each [referring to the co-ordination of a phrase and the sense] is particular; it has its uses and its characteristic environment". He thus rejects the common teaching idea of presenting PVs as a 'featureless list'. With information on the phraseologies of each PV, the PVs could be presented more meaningfully.

Incorporating phraseologies into PV teaching fits well with theories of lexical knowledge. What it means to know a lexical item has been sketched by researchers like Nation (2001), who points out that dimensions of vocabulary knowledge include, for example, associations, collocations and constraints and so on (see Section 3.2.2). Read (2004) also outlines that producing a word successfully involves 'precision of meaning', 'comprehensive word knowledge' and 'network knowledge'. A language user has to know the clear content of the word sense, with all the relevant elements such as collocations, syntax and pragmatic requirements, and be able to build networks with other words in the mental lexicon. Depth of lexical knowledge is considered vital for learners. If an advanced goal is pursued, that is the achievement

of nativeness, learners are further expected to possess knowledge of ‘typicality’: they need to be aware of what is most frequent in a language (Stubbs, 2001).

The information of lexical knowledge and typicality mentioned above can be revealed to students by comparing a native speaker corpus and a learner corpus, as this present study has done. Consulting a native speaker corpus can be used to increase the amount of information available to learners, and so facilitate the development of in-depth lexical knowledge, and on the other hand, investigating a learner corpus can be used to assess the learners’ current state of lexical knowledge.

In practice, students’ lexical knowledge can be improved through using the DDL approach, which makes great use of corpora (see Section 3.3.1). Students can discover the relevant information either by exploring the concordance lines themselves or by teachers’ guidance. For easily-observed patterns, students are encouraged to find the answers on their own, while for the phraseological units which are flexible or abstract; teachers may offer help where appropriate, as this present study has shown the existence of these flexible or abstract phraseological units, indicating that they are not easy to perceive but are crucial to a lexical item’s usage.

To conclude the pedagogical recommendations, I have not found particular reasons to list PVs as distinguished units in the syllabuses/textbooks. However, PVs

should be taught systematically by introducing PVs with the same particle, using the cognitive approach. Once the particle group is decided, the teacher can start to select appropriate PVs for teaching based on students' proficiency levels, with the assistance of the two-axis model. The selected PVs can be joined with other vocabularies so that the teacher can create texts of a specific topic, genre or notion, with which these PVs are found to be preferably used. If any of the selected PVs have synonymous counterparts, the different usages of the near-synonym can be deliberately included in the texts, so that their disparities are made visible to students. After reading these made-up texts, students will be asked to guess the meanings of the PVs from the context. The next step is to use the DDL approach, which presents to students the concordance examples retrieved from corpora to reveal the typical phraseological patterns of each PV. The key to teaching lies in training learners to successfully perceive the phraseologies and to reproduce them accurately in new contexts.

## **9.5 Summary**

The summative findings reveal the Chinese learners' idiosyncratic features. Four distributional differences are noted, of which some are in accordance with existing studies (e.g. the high TTRs of LOCNESS) and some display evidences counter to

what has been found before (e.g. the greater frequencies of PVs in CLEC). Two parameters of phrasal verb features have also been proposed and discussed, pinpointing the significance of idiomaticity and collocation strength in either productive or receptive tasks. The comparison of synonymous PVs and single-word verbs suggests that identifying their phraseological units helps clarify their meanings and usages. The Chinese learners' idiosyncrasies shown at all phraseological levels suggest that more attention should be drawn to them. These idiosyncrasies can be traced to the differences between text types (topics/genres/registers) and the first language transfer, and accounted for by the theories of L1-L2 conceptualisation. Several innovative pedagogical recommendations were made based on the findings of this thesis, including the classing, selecting, ordering and contrasting of PVs and the consideration of their phraseologies.

# **Chapter10: CONCLUSION**

## **10.1 Introduction**

This final chapter will first review the thesis and discuss the main conclusions and contributions of this work (Section 10.2). It will then go on to assess the limitations of this thesis (Section 10.3). For future avenues, some caveats will be given in Section 10.4.1, with the directions of future research discussed in Section 10.4.2. The possible challenges of applying the results in practice are presented in Section 10.4.3. The final section is a brief one which draws together the dominant ideas of the bulk of the thesis.

## **10.2 Review of the thesis**

This thesis set out to explore the usage differences of PVs between the Chinese learner English and the native English (the main research question, see Section 1.5), by looking at the phraseological units of PVs, including collocates, semantic fields, semantic sequences and prosody. Along with the analyses of phraseologies, this thesis has also investigated the distributions of PVs and the contrasts of near-synonyms with PVs, and has proposed an innovative two-dimensional model to account for the categorisation of PVs (see also Section 1.5, for research questions).

A Chinese learner corpus (CLEC) was compared with a native English corpus

(LOCNESS), and a reference corpus (BoE) was consulted where necessary. Phrasal verbs with five different particles were identified using appropriate computer programmes. This study not only surveyed the frequencies of PVs of the five particle groups and calculated the type-token ratios, but also presented in detail a series of case studies of PVs which appear to be worth further research, based on the quantitative data.

From Chapters 6 to 9, I have analysed the frequencies and the usages of PVs in five different particle groups. Each chapter focused on PVs with a specific particle and a particular issue of theory. The distributional research and case studies in this current work have successfully identified the differences of phrasal verb behaviours between the two languages. The key findings of this thesis have been outlined and discussed in Chapter 9, and are summarised in a list as below:

1. Chapter 6 first investigated the frequencies and TTR of the PVs with the particle UP, and then carried out case studies of five PVs. The problem of ambiguity was noted in the literature and the solution of the two-dimensional model was put forward. It was found that the PVs with UP occurred much more frequently in CLEC but the Chinese learner language appeared to have lower type-token ratios, which indicate that the Chinese learners used fewer phrasal verb types but used each one more frequently than the native students. The analyses of the five PVs revealed that some idiosyncratic collocates and

semantic concepts were used by the Chinese learners alone. This analysis suggested in turn that most of the literature on PVs has confused two parameters of PVs: degrees of idiomaticity and restriction strength. A two-dimensional model thus was proposed to solve this problem.

2. Chapter 7 scrutinised the frequencies and TTR of PVs with OUT. It then focused on two specific PVs (CARRY OUT and FIND OUT), followed by a comparison of FIND OUT and the synonymous verb FIND. The distributional results are like those found in Chapter 6. In the two case studies, more idiosyncratic collocates and semantic concepts were witnessed. The comparison of FIND OUT and FIND showed that their usages, while have some overlaps, are actually not identical to a certain extent.
3. Chapter 8 explored the PVs of three particle groups ON, ABOUT and DOWN. The Chinese learners were found to use more PVs in one of the particle group (the DOWN group), and have lower TTRs in all groups except the ABOUT group. The case studies of these three particle groups also pointed to that the Chinese learners' usages of PVs differ from the native usages in terms of phraseological units.

A set of implications thus could be summarised based on these findings:

- (a) A general conclusion is that the PVs are used in different ways by the Chinese learners in comparison with English writers, both quantitatively and qualitatively. The frequency data and the usage patterns of PVs have demonstrated that the two groups of writers indeed show their own characteristics in using PVs.

(b) More specifically, the Chinese learners' language can be characterised by having less diversity and more repetition of phrasal verbs, and by the preference for literal PVs over idiomatic ones. These characteristics are in line with the descriptions of PVs in the literature.

(c) Learner-specific uses of PVs were found at all the phraseological levels, where L1 influence was found to be involved. This adds more evidences to support Ellis' (2008: 8) conclusion that "transfer affects L2 phraseology at numerous levels", which itself summarises a range of previous research.

A more significant finding is that, in the Chinese learners' uses of PVs, not only words but also concepts can deviate from the native uses. This finding has the potential to suggest to future researchers a new direction of studying learner language, as it is apparent that learner language can deviate from the native language both explicitly (words) and implicitly (concepts). As such, attention should also be drawn to the deviations of concepts, which usually can be identified by flexible words or abstraction of meanings. As has been insightfully proposed by Kilgarriff (1997:108), word senses are "abstractions over clusters of word usages". Seen from this perspective, the meaning of a word has to be decided by its context and usage. Kennedy (2008:38) clearly criticises the existing studies for largely addressing the

forms of phraseologies: they are conducted on the exact words which constitute the composition, while the abstractions of these words, such as semantic relationships, have been relegated to the periphery. In the same vein, Shei (2008:69) calls on teachers to draw students' attention to the 'prototypical' ELU (Extended Lexical Unit in Stubbs' (2002) terms), which is similar to the phraseological units analysed in this thesis. My results provide support to these ideas and underline the overriding need to cover these flexible and hidden units.

(d) Another important conclusion is that the phraseological behaviours of PVs are individualised. The individualisation can be noticed at all of the levels which constitute phraseological units, including collocation, colligation, semantic preference, semantic prosody, and so on. Each node word (in this case the phrasal verb) manifests different phraseological characteristics. Some PVs have salient features at one level, while others have prominent phenomena at other levels. For example, one phrasal verb may have a predominant semantic preference, whilst other PVs may have other noticeable features such as semantic sequences. In other words, the phraseological behaviours of each phrasal verb are likely to vary. Not every phraseological unit is discovered from each selected PV. The fact that each PV has individualised phraseologies conforms to the observation from Partington (1998:27) that "every

lexical item in the language has its own individual and unique pattern of behaviour”.

The distinct patterns of each individual PV imply that the phraseologies of a PV can characterise its usage, and the usage of a PV represents its ‘identity’.

(e) It is useful to categorise PVs with a two-dimensional model which takes into account the degrees of idiomaticity and restriction strength. Categorising PVs with these two parameters can clearly show the features of different groups of PVs. Furthermore, this model has pedagogical value that it serves to estimate the difficulty degrees of PVs, for both teachers and learners. Moreover, for better learning, we should be on alert that the synonymous single-word verb and the PV act differently, as has been informed by this study.

In light of these conclusions, this study has made contributions in four respects: First, it has widened our understanding of learner languages and has shown that the concept of “phraseological units” are a good means to define the usages of PVs or other lexical items, and this harmonises with the theories of vocabulary knowledge. Second, it has extended the focus in phraseology studies from fixed sequences of words to flexible combinations of concepts (and words), allowing for a full picture of usages to emerge. Third, it has clarified the complicated properties of PVs by separating idiomaticity and restriction strength, two crucial factors which have tended

to be confused, in a newly created two-dimensional model. Fourth, it has thrown light on the teaching of PVs to L2 students, thus specific and practical advices were able to be made. Besides these primary ones, a secondary contribution is that the study has uncovered collections of important PVs (viz. PVs of high frequency) in native English (LOCNESS) and the Chinese learner language (CLEC) and recorded the occurrences and TTRs (Section 9.2.1). The frequent PVs identified were used as the data pool, from which interesting items were selected for further analysis. These frequent PVs not only prepare the ground for future studies but also pave the way for establishing appropriate learning materials.

### **10.3 Limitations**

The methods and analyses have successfully identified the results of the problems which this thesis aimed to solve. However, the limitations encountered in this study must be highlighted.

One of the limitations concerns the materials, that is, the corpora and inclusion of targets. With respect to the corpora, a variable which needs improvement is the size of the NS and NNS corpora. Given the infrequency of phrasal verbs, much larger corpora would likely yield more occurrences of PVs and would allow us to have more

confidence in the results. With respect to the inclusion of targets, the reliability of this study could also be much enhanced if the genre/register/topic types of the corpora were strictly controlled, as this study has found they affect the occurrences of PVs. In hindsight, a consistency of genre/register types of text and the topics of essays in LOCNESS and CLEC would no doubt render the corpora more comparable. Unfortunately, corpora that are consistent in this respect do not yet exist, and it would not have been feasible for me to collect such corpora of a satisfactory size. Furthermore, although the analysis of this study is sufficient for interpreting the behaviours of PVs, it covered not all PVs but only some of them with different particles. It would be worthwhile extending the investigation to as many varieties of PVs as possible, because they may provide more richness of data.

A second limitation relates to the use of reference corpora. In the case of this present research, I consulted BoE wherever the linguistic items were not presented in LOCNESS because LOCNESS is rather small. The problem is that the results from the two English corpora are not always identical: a phenomenon found in one corpus may not occur in the other. This is predictable, as the properties (e.g. tokens, genres, writer backgrounds) of these two corpora are not identical. The inconsistencies between BoE and LOCNESS certainly cast doubt on the validity of the corpora

comparison. Two corpora can only be said to be completely comparable if they are rigidly compiled in every way. Many cases found in BoE cannot be attested to in LOCNESS due to its small size, but although the BoE is massive, its components are not in accord with CLEC. Therefore a larger native corpus compiled following the criteria of a learner corpus is desirable in the near future.

A third limitation regards the extraction method. One problem found from the application of computer programs in retrieving language targets is the handling of inflected word forms. During the extraction process, it was revealed that not all the collocates could be captured by WordSmith4. There are two reasons for the loss of collocates: First, WordSmith4 does not have a lemmatising function, so different wordforms belonging to the same lemma are treated as different types. Secondly, the WordSmith 'Collocate' program lists the collocates in each position (Left 1, Left 2, etc.) separately, so, for example, if a word appears 10 times in L1 and 5 times in L2, there is no automatic way of telling that the word occurs as a Left collocate 15 times.

For these reasons, frequencies of collocates are likely to be mistaken, and will probably be underestimated, unless all the collocates in all positions are carefully checked manually. As seen in Section 4.2, Altenberg and Granger (2001:185) point out that the part-of-speech of each collocate is left unknown in the WordSmith tool as

well. They are also critical of the fact that many identified collocates are not so-called ‘constructional collocates’, that is, collocates which bear at least some semantical relation (e.g. *time+goes on*). Therefore if this thesis is to be generalised to other linguistic targets, a heavy workload of manual analysis will be required.

## **10.4 Future avenues**

### **10.4.1 A word of caution for future research**

Some caveats to future research have been noted in this study. A caveat arose from the investigation process of this study is the danger of examining PVs using only a quantitative method. As noted (see Chapter 2), most phrasal verb studies have adopted straightforward frequency of occurrence as the single means to analyse them. Nevertheless, the fact that many PVs have multiple senses brings into doubt the adequacy of such a method of data interpretation. Having looked into the behaviours of each high-frequency phrasal verb, this thesis suggests that each occurrence of different senses should be analysed independently and counted separately. However, determining the senses of PVs by co-texts is a formidable task at present, since corpus studies often deal with a great amount of data. Unfortunately the software that I have used does not contribute much to solving this problem and this task still relies heavily on human analysts. Therefore an important task for future research is to prioritise the development of automatic sense-disambiguation approaches, while at the same time

recognising that human involvement is still a necessity.

Another caveat relates to the categorisation of semantic fields. When the collocations were probed, it was up to the researcher to decide the exact semantic fields; however, the semantic fields can be described focusing on different facets. Take the collocates of FIND OUT, for example: the collocated nouns were grouped according to their meanings such as solutions (e.g. *answer*) or facts (e.g. *truth*). This is not the only way to classify these collocates: they can also be divided into something which is hidden (e.g. *secret*) or something which is new (e.g. *information*). The determination of semantic fields is largely dependent on the criteria the researcher chooses. Moreover, these subordinate fields can also be covered by a superordinate concept that they are all referring to something unknown, unavailable to the speaker, and thus waiting to be 'found out'. The fact that semantic fields can be depicted in many ways will pose problems once the semantic fields of collocation are presented to foreign language learners. The extent of delimitation will affect how the learners associate and combine words and concepts and how their L2 lexical networks are arranged.

## **10.4.2 Directions for future research**

The findings of this study will lead future research to two main directions, with respect to software development and research of phraseologies.

### ***10.4.2.1 Software development***

The investigation process has highlighted the urgent need to develop appropriate software and techniques to deal with phraseological units at higher levels. Corpora are useful means that contribute to describing languages from a new perspective. The nature of a corpus as a collection of texts has great consequences for methods of investigation. Because such a large amount of data can be effectively processed by computers, dependence on the ways in which computers cope with data is apparent. The data format must be typewritten for the computers to read and retrieve it. As such, corpus programmes will be good at presenting patterns ‘discernible’ to computers (the advantage is that these patterns will have been previously unnoticeable to the human eye), that is, regularities in the concordance. Some of the phraseological concepts cannot be discovered by computer programmes, which react only to tangible words. What cannot be easily captured is the regularities beyond words, such as semantic or discoursal constraints, which have been found crucial in this study. Once these units can be identified by computers, a substantial volume of results can then be obtained more efficiently and our language learning instruction can be improved as well.

#### ***10.4.2.2 Research on phraseologies***

This thesis pioneered research in the phraseologies of phrasal verbs, in response to the urgent call made by Waibel (2007), and has validated the phraseologies of LL that feature the Chinese learner idiosyncrasy. The results of this study can inform further research of three new perspectives:

First, the phraseological units are better regarded as an integrated whole rather than separate concepts. These units have interactions with each other (all of them work together to form meanings and usages), and their boundaries are sometimes hard to demarcate. They are, in turn, better tackled together rather than treated in isolation. The study of phraseology may become more useful if all potential phraseological units are taken into account.

Second, this thesis has also drawn attention from formulae to strings which have flexible fragments or implicit connotations, such as semantic sequences and semantic prosody. I looked at the behaviours of PVs through the lens of phraseologies, from which many interesting findings were uncovered, especially those which can only be detected by human eyes. It will be worthwhile to explore L2 learners' phraseological patterns, which allow more flexibility or variability than formulae.

A final reminder to future studies is to pay great attention to the variables of text types such as topic/genre/register, and the considerable influence of first languages.

As discussed in Section 9.3, if comparisons of native and learner corpora are to be drawn in the future, these factors must be carefully controlled in order to prevent skewed results.

This thesis has also thrown up some questions which are worthy of further consideration. For example, researchers may want to ask whether longer phraseological units are psycholinguistically valid (see the theory of lexical priming in Hoey (2005)). Or they may like to consider whether the L1 phraseologies reflect the learner-specific patterns empirically. Another interesting area is to explore the development of PVs by taking into account the learners' proficiency levels. All of these questions can serve as points of departure for future research.

### **10.4.3 The challenges of learning phraseologies**

The ultimate goal of this current study is to apply the findings to improve English learning and teaching. In this study, it is advocated that the concept of 'phraseological units' to be utilised to account for the usages of words, through the assistance of corpora. However, there may not be a straight application, and a discussion of the possible challenges seems to be in order at this point, because it is too important to be left.

First, there is a gap to bridge between corpus data as product and the process of language production. In other words, the results produced from corpora may not correspond to the text/speech production process. The analysis of corpora is basically ‘bottom-up’, in that the large collection of data is filtered and reduced to samples of concordance lines which allow patterns to emerge. On the contrary, the process by which a language user produces text or speech is ‘top-down’, as described in Denes and Pinson (1963:3) (cited in (Gleason & Ratner, 1998:311)). A speaker/writer:

*“has to [...] arrange his [sic] thoughts, decide what he wants to say, and put what he wants to say into linguistic form [...] by selecting the right words and phrases to express its meaning, and by placing these words in the correct order required by the grammatical rules of the language .”*

Frank et al. (2012), drawing on evidence from neuroscience, computational and psycholinguistic studies, argue that a simple linear combination of language items (e.g. words, phrases) can more adequately model the operation of how language is used. In other words, language use is more likely to be ‘sequential’ than ‘hierarchical’. Although the underlying principle of this conclusion is similar to corpus studies stating that language use is linear, the workings of transforming an intention into concrete language in the form of text or speech in the language user’s mind have not

been completely clear. Thus it is logical to infer that corpus findings do not correspond to language production because the former is a product while the latter is a process. If the results of corpus studies do not reflect the actual way in which language is produced by a writer or speaker in practice, this will cause serious problems in applying corpus data to L2 learning.

Using corpora in the classroom or adopting corpus results for teaching and learning has been advocated by many researchers (see Chapter 3), and the corpus is truly a powerful device which can assist learning. The application of corpora, however, has to be undertaken with caution. We could conceive that when a L2 learner attempts to convey a message, they will have to choose appropriate words and put them in proper sequences. It seems efficient to show the learner patterns from the corpus, so that they can engage in the adequate selection and ordering of the words at their disposal. Unfortunately, in reality the learner will be overwhelmed by the substantial amount of data, as pointed out by Cook (1998: 61), “the description of English which emerges from corpus analysis [...] is dauntingly complex and particular”. The overload of data provided by corpora will inevitably become an impenetrable barrier for L2 learners.

The application of phraseological units has also been questioned by Hunston

(2007), who argues that semantic prosody has more observational than predictive value. Likewise, she also warns that the function of semantic sequences is descriptive rather than prescriptive (Hunston, 2009:151), thus semantic sequence may have less value than expected to learners because the learners do not “distinguish between correct and incorrect sequences” (Hunston, 2009:153). Nonetheless, she affirms that the phraseologies of learner data can be used for syllabus design (Hunston, 2009:151). Indeed, for these phraseological units which do not contain only fixed words, their combinations are tendencies but not rules. There is no right or wrong in using them, and learners apparently cannot ‘learn’ them by memorisation. Even so, presenting this information to learners can still help to raise their awareness of the differences between the LL and native conventions. Such awareness can equip them to avoid learner-specific phraseologies and follow sets of habits in English (to concatenate words like English writers or ultimately even to think like them).

## **10.5 Concluding remarks**

This study has explored phrasal verb usages in the Chinese English learners’ language by comparing it to native writings, using a contextual approach which involves multiple phraseological notions such as collocates, semantic preferences, semantic

prosody and semantic sequences. It has also sought to find alternative solutions for improving learners' knowledge about phrasal verbs, with respect to categorising phrasal verbs with their defining criteria and distinguishing synonyms by describing different behaviours.

From the overview of this thesis, we could conclude that the Chinese learners' use of phrasal verbs exhibits many learner-specific features. The Chinese learners' uses of PVs were found to be more prevalent (i.e. to have more tokens) but less heterogeneous (i.e. to have fewer varieties of PV types). Case studies further showed that each phrasal verb has phraseologies of its own, and therefore the Chinese learners of phrasal verbs will face not only syntactic and semantic complexities, but also varied behaviours of individual phrasal verbs. The Chinese learner language idiosyncrasies are manifested in these phraseologies. Such learner language idiosyncrasies are regular and systematic, suggesting that the lexical items investigated here are differently structured or linked in the learners' L2 lexicon as compared to the natives'. Since the first language has a great influence on learners' L2 performance, it is crucial to teach learners the conceptual differences between L1 and L2, which can be reflected by comparing the L1-L2 phraseological differences. In this respect, this thesis suggests that studies of phraseology should use a more flexible

approach similar to the one adopted in this study, because such an approach takes all phraseological units into account, and can allow more interesting phenomena to emerge.

I hope that this thesis promotes the importance of raising learners' awareness of phraseologies, since knowledge of these phraseological units is indispensable in learning a language. Also, I hope that the approach employed in this thesis will bring to light what prior studies did not account for, and help advance our understanding of learner languages.



**Appendix A: Frequency list of 'Verb+UP'**

Verb Type	CLEC		LOCNESS	
	abs.	rel.	abs.	rel.
act	1	0.93	0	0
add	7	6.54	0	0
back	1	0.93	8	24.67
beat	3	2.8	2	6.17
bind	1	0.93	0	0
block	1	0.93	0	0
boil	0	0	2	6.17
bottle	0	0	1	3.08
break	8	7.47	2	6.17
brighten	0	0	1	3.08
bring	15	14.01	38	117.17
buckle	0	0	1	3.08
build	56	52.31	10	30.84
burn	1	0.93	0	0
buy	0	0	1	3.08
call	3	2.8	0	0
catch	35	32.69	4	12.33
check	2	1.87	1	3.08
cheer	6	5.6	0	0
chop	1	0.93	0	0
clean	4	3.74	3	9.25
clear	6	5.6	3	9.25
clog	0	0	1	3.08
come	32	29.89	12	37
cover	3	2.8	2	6.17
crop	3	2.8	0	0

cut	2	1.87	1	3.08
dig	0	0	1	3.08
divvy	0	0	1	3.08
drag	0	0	2	6.17
draw	16	14.94	9	27.75
dress	4	3.74	1	3.08
drink	1	0.93	0	0
dry	4	3.74	0	0
eat	9	8.41	0	0
end	13	12.14	30	92.51
even	0	0	1	3.08
face	7	6.54	0	0
fill	5	4.67	1	3.08
fix	1	0.93	0	0
flare	1	0.93	1	3.08
follow	4	3.74	2	6.17
free	0	0	1	3.08
freshen	2	1.87	0	0
fuel	1	0.93	0	0
get	161	150.38	4	12.33
give	157	146.65	30	92.51
go	22	20.55	1	3.08
grow	88	82.2	30	92.51
hang	20	18.68	1	3.08
heat	0	0	1	3.08
help	2	1.87	0	0
hold	21	19.62	7	21.58
hook	0	0	1	3.08
hurry	10	9.34	0	0
join	0	0	1	3.08
jump	7	6.54	0	0

keep	49	45.77	3	9.25
lead	1	0.93	0	0
leave	0	0	1	3.08
lift	4	3.74	0	0
light	3	2.8	1	3.08
line	1	0.93	4	12.33
link	0	0	2	6.17
look	50	46.7	0	0
make	106	99.01	25	77.09
measure	0	0	1	3.08
meet	0	0	1	3.08
mix	1	0.93	1	3.08
move	2	1.87	0	0
open	3	2.8	7	21.58
pack	1	0.93	0	0
pair	0	0	1	3.08
pass	2	1.87	1	3.08
pay	3	2.8	0	0
pick	50	46.7	12	37
pile	4	3.74	0	0
pop	0	0	1	3.08
pull	5	4.67	0	0
push	0	0	1	3.08
put	25	23.35	5	15.42
queue	1	0.93	0	0
ring	14	13.08	0	0
rise	27	25.22	0	0
roll	2	1.87	0	0
round	0	0	1	3.08
run	5	4.67	11	33.92
save	12	11.21	1	3.08

screw	0	0	1	3.08
sell	1	0.93	1	3.08
send	1	0.93	0	0
set	100	93.41	17	52.42
shake	1	0.93	1	3.08
show	3	2.8	2	6.17
shut	1	0.93	0	0
sign	3	2.8	0	0
sit	2	1.87	0	0
smash	1	0.93	0	0
snap	1	0.93	0	0
speak	0	0	2	6.17
speed	9	8.41	4	12.33
spring	2	1.87	2	6.17
stand	43	40.16	2	6.17
start	2	1.87	0	0
stay	11	10.27	0	0
step	1	0.93	2	6.17
stick	2	0.87	0	0
stir	0	0	1	3.08
straighten	1	0.93	0	0
suck	0	0	1	3.08
sum	27	25.22	3	9.25
swallow	1	0.93	0	0
take	101	94.34	12	37
tear	1	0.93	0	0
throw	1	0.93	1	3.08
tie	2	1.87	1	3.08
tighten	1	0.93	0	0
trip	1	0.93	0	0
turn	7	6.54	1	3.08

use	115	107.42	0	0
wake	58	54.18	5	15.42
walk	3	2.8	0	0
warm	1	0.93	0	0
wash	4	3.74	0	0
weigh	0	0	1	3.08
whip	0	0	1	3.08
whoop	1	0.93	0	0
wind	0	0	1	3.08
work	2	1.87	1	3.08
wrap	1	0.93	1	3.08
TOTAL	1630	1521.42	363	1119.19

### Appendix B: Frequency list of 'Verb+OUT'

Verb	CLEC		LOCNESS	
	abs.	rel.	abs.	rel.
act	5	4.7	5	15.4
assure	1	0.9		0.0
back		0.0	2	6.2
battle		0.0	1	3.1
bear		0.0	2	6.2
block		0.0	1	3.1
blow	3	2.8	7	21.6
blurt	1	0.9		0.0
break	31	29.0	2	6.2
bring	14	13.1	10	30.8
build	2	1.9		0.0
bump	1	0.9		0.0
burn	1	0.9		0.0
burst	6	5.6		0.0
buy	1	0.9		0.0
call	38	35.5	2	6.2
cancel	1	0.9		0.0
carry	100	96.2	65	200.4
catch	1	0.9		0.0
chase	1	0.9	1	3.1
cheat	1	0.9		0.0
check	4	3.7	2	6.2
chew		0.0	1	3.1
chill		0.0	1	3.1
churn		0.0	1	3.1
clear	1	0.9	1	3.1
come	49	45.8	10	30.8
contract	1	0.9	1	3.1
count	1	0.9		0.0
crash	2	1.9		0.0
crowd	1	0.9		0.0
cry	24	22.4	3	9.3
cut	1	0.9		0.0
dash	4	3.7		0.0
develop	1	0.9		0.0
die	15	14.0		0.0
do	1	0.9	1	3.1
dope	2	1.9		0.0
drag	1	0.9		0.0
drain	1	0.9		0.0
draw	2	1.9		0.0
drill	1	0.9		0.0
drink	1	0.9		0.0
drip	1	0.9		0.0
drive	4	3.7	3	9.3
drop	8	7.5	4	12.3
ease	1	0.9		0.0
eat	1	0.9		0.0

fall	3	2.8		0.0
fight		0.0	1	3.1
figure	5	4.7	2	6.2
fill		0.0	2	6.2
find	120	109.3	25	77.1
finish	2	1.9		0.0
flare		0.0	1	3.1
flash	1	0.9		0.0
flee	1	0.9		0.0
float	1	0.9		0.0
flood	1	0.9		0.0
flow	1	0.9		0.0
fly	1	0.9		0.0
force	2	1.9		0.0
fork		0.0	1	3.1
get	50	46.7	14	43.2
give	16	14.9	2	6.2
go	301	281.2	17	52.4
gouge		0.0	1	3.1
grow	1	0.9	1	3.1
guess	2	1.9		0.0
hand	6	5.6	2	6.2
hang	4	3.7		0.0
heat	1	0.9		0.0
help	10	9.3	1	3.1
hit		0.0	2	6.2
hold		0.0	1	3.1
hunt		0.0	1	3.1
hurry	16	14.9		0.0
imagine	1	0.9		0.0
jump	57	53.2		0.0
keep	12	11.2	4	12.3
kick		0.0	3	9.3
kill	1	0.9		0.0
knock		0.0	4	12.3
lash		0.0	1	3.1
lay	7	6.5	3	9.3
leap	1	0.9		0.0
leave	8	7.5	5	15.4
let	20	18.7	1	3.1
lift		0.0	1	3.1
litter	1	0.9		0.0
live	2	1.9	2	6.2
look	21	19.6		0.0
loose		0.0	2	6.2
lose		0.0	7	21.6
make	18	16.8	6	18.5
map	1	0.9		0.0
march	1	0.9		0.0

mellow		0.0	1	3.1
miss	2	1.9	4	12.3
move	6	5.6	3	9.3
open		0.0	1	3.1
pass	1	0.9		0.0
pay		0.0	2	6.2
peep	1	0.9		0.0
pick	11	10.3	3	9.3
place	1	0.9		0.0
plan		0.0	1	3.1
play		0.0	2	6.2
point	30	28.0	41	126.4
pour	7	6.5		0.0
prevent	1	0.9		0.0
print		0.0	1	3.1
produce	1	0.9		0.0
pull	2	1.9	5	15.4
push	2	1.9	1	3.1
put	60	56.0	2	6.2
reach	2	1.9	1	3.1
read		0.0	1	3.1
reject	1	0.9		0.0
remove		0.0	1	3.1
reveal	2	1.9		0.0
rid	1	0.9		0.0
root	2	1.9		0.0
rule	4	3.7	1	3.1
run	43	40.2	7	21.6
rush	32	29.9	2	6.2
say	2	1.9		0.0
scream	11	10.3	1	3.1
scout		0.0	1	3.1
see	2	1.9		0.0
seek		0.0	3	9.3
sell	7	6.5	2	6.2
send	18	16.8	1	3.1
serve	1	0.9		0.0
set	12	11.2	12	37.0
share		0.0	1	3.1
ship		0.0	1	3.1
shout	1	0.9		0.0
show	1	0.9		0.0
sift	1	0.9		0.0
single	3	2.8	3	9.3
sit		0.0	1	3.1
snap		0.0	1	3.1
snuff		0.0	1	3.1
sob	1	0.9		0.0
sort	3	2.8	4	12.3
spat	1	0.9		0.0

speak	20	18.7	3	9.3
spend	1	0.9		0.0
splash	1	0.9		0.0
spread	3	2.8	2	6.2
spring		0.0	1	3.1
stamp		0.0	1	3.1
stand	4	3.7	3	9.3
stare		0.0	1	3.1
start	2	1.9	9	27.8
stay	1	0.9		0.0
step	13	12.1	1	3.1
stick	4	3.7		0.0
straighten	1	0.9		0.0
stress		0.0	1	3.1
stretch	4	3.7	1	3.1
strive	1	0.9		0.0
swarm	1	0.9		0.0
sweep	1	0.9		0.0
take	57	53.2	9	27.8
talk		0.0	2	6.2
tell	1	0.9		0.0
think	20	18.7	6	18.5
throw	3	2.8	6	18.5
toss		0.0	1	3.1
train		0.0	1	3.1
travel	1	0.9		0.0
try	5	4.7		0.0
tune		0.0	1	3.1
turn	43	40.2	11	33.9
type	1	0.9		0.0
use	16	14.9		0.0
walk	17	15.9	2	6.2
want		0.0	2	6.2
wash	1	0.9		0.0
watch		0.0	3	9.3
wear	19	17.7	1	3.1
wedge	1	0.9		0.0
weed		0.0	1	3.1
win		0.0	1	3.1
wipe	3	2.8	6	18.5
wonder	1	0.9		0.0
work	48	44.8	7	21.6
<b>TOTAL</b>	<b>1603</b>	<b>1497.3</b>	<b>434</b>	<b>1338.3</b>



## Appendix C: Frequency lists of 'Verb+ON', 'Verb+ABOUT', 'Verb+DOWN'

*Table 1: Frequency list of 'Verb+on'*

Verb Type	CLEC		LOCNESS	
	abs.	rel.	abs.	rel.
act	2	1.87	0	0
base	1	0.93	0	0
bring	0	0	2	6.17
build	0	0	1	3.08
call	1	0.93	0	0
carry	14	13.08	12	37
catch	1	0.93	0	0
cling	0	0	2	6.17
come	5	4.67	1	3.08
decide	0	0	1	3.08
draw	1	0.93	0	0
follow	0	0	1	3.08
focus	1	0.93	3	9.25
get	41	38.3	5	15.42
go	114	106.48	55	169.59
hand	1	0.93	0	0
hang	9	8.41	2	6.17
hold	2	1.87	4	12.33
insist	4	3.74	1	3.08
jump	1	0.93	0	0
keep	41	38.3	1	3.08
lay	1	0.93	0	0
linger	0	0	1	3.08
live	22	20.55	4	12.33
look	5	4.67	0	0
move	6	5.6	2	6.17

pass	3	2.8	9	27.75
pick	0	0	1	3.08
play	3	2.8	0	0
push	1	0.93	0	0
put	27	25.22	2	6.17
rely	2	1.87	0	0
report	0	0	1	3.08
run	2	1.87	1	3.08
sell	1	0.93	0	0
sign	0	0	1	3.08
spur	0	0	1	3.08
stay	0	0	1	3.08
stick	1	0.93	0	0
switch	1	0.93	0	0
take	31	28.96	33	101.76
touch	0	0	1	3.08
try	1	0.93	0	0
turn	7	6.54	1	3.08
urge	0	0	1	3.08
walk	2	1.87	1	3.08
work	2	1.87	0	0
TOTAL	357	353.07	152	468.64

Table 2: Distribution of 'Verb+about' in alphabetical order

Verb Type	CLEC		LOCNESS	
	abs.	rel.	abs.	rel.
bounce	0	0	1	3.08
bring	59	55.11	27	83.26
come	0	0	9	27.75
concern	1	0.93	0	0
doubt	1	0.93	0	0
go	1	0.93	0	0
hang	1	0.93	0	0
hustle	1	0.93	0	0
inform	0	0	1	3.08
know	4	3.74	1	3.08
learn	1	0.93	0	0
move	1	0.93	0	0
ponder	1	0.93	0	0
read	2	1.87	2	6.17
report	0	0	1	3.08
run	0	0	1	3.08
say	1	0.93	0	0
set	3	2.8	0	0
slouch	1	0.93	0	0
swim	1	0.93	0	0
think	1	0.93	0	0
toss	1	0.93	0	0
wonder	3	2.8	0	0
worry	7	6.54	1	3.08
<b>TOTAL</b>	<b>91</b>	<b>85</b>	<b>44</b>	<b>135.68</b>

Table 3: Distribution of 'Verb+down' in alphabetical order

Verb Type	CLEC		LOCNESS	
	abs.	rel.	abs.	rel.
beat	2	1.87	0	0
bend	1	0.93	0	0
blow	1	0.93	0	0
boil	1	0.93	0	0
break	12	11.21	16	49.34
bring	0	0	6	18.5
calm	4	3.74	1	3.08
chase	0	0	3	9.25
climb	1	0.93	0	0
close	12	11.21	1	3.08
come	10	9.34	6	18.5
cool	4	3.74	0	0
crack	5	4.67	1	3.08
crush	2	1.87	0	0
cut	24	22.42	8	24.67
die	1	0.93	2	6.17
drop	9	8.41	0	0
fall	45	42.03	1	3.08
flow	1	0.93	0	0
flutter	4	3.74	0	0
fly	2	1.87	0	0
get	15	14.01	3	9.25
go	16	14.94	9	27.75
grind	0	0	1	3.08
gulp	1	0.93	0	0
hand	7	6.54	1	3.08
hang	2	1.87	0	0
head	0	0	1	3.08

hit	2	1.87	0	0
hold	3	2.8	2	6.17
hunt	1	0.93	1	3.08
jot	1	0.93	0	0
jump	3	2.8	0	0
keep	0	0	3	9.25
kneel	2	1.87	0	0
knock	65	60.71	1	3.08
lay	11	10.27	3	9.25
let	0	0	2	6.17
lie	8	7.47	1	3.08
look	27	25.22	8	24.67
march	1	0.93	0	0
mark	3	2.8	0	0
move	1	0.93	0	0
narrow	2	1.87	0	0
note	2	1.87	0	0
pass	0	0	3	9.25
place	0	0	1	3.08
press	1	0.93	0	0
pull	2	1.87	0	0
push	1	1.87	1	3.08
put	13	12.14	2	6.17
rain	1	0.93	0	0
roll	0	0	1	3.08
run	3	2.8	0	0
send	0	0	1	3.08
set	3	2.8	1	3.08
settle	4	3.74	0	0
shake	1	0.93	0	0
shoot	2	1.87	1	3.08

shut	2	1.87	0	0
sit	46	42.97	2	6.17
skid	1	0.93	0	0
slide	1	0.93	0	0
slip	1	0.93	0	0
slow	8	7.47	6	18.5
squat	1	0.93	0	0
stab	1	0.93	0	0
stand	0	0	1	3.08
stay	2	1.87	0	0
step	2	1.87	1	3.08
strike	2	1.87	0	0
swoop	0	0	1	3.08
take	2	1.87	2	6.17
tear	1	0.93	0	0
throw	2	1.87	0	0
tone	0	0	1	3.08
track	0	0	1	3.08
trickle	1	0.93	0	0
turn	1	0.93	0	0
wear	0	0	1	3.08
weigh	0	0	1	3.08
write	28	26.15	3	9.25
TOTAL	447	417.52	112	345.35

APPENDIX D

Table1: Senses of 'FIND' in dictionaries

No.	Collins	Cambridge	Merriam
1 [VERB] V n, V n n, also V n for n	<p>If you <b>find</b> someone or something, you see them or learn where they are.</p> <p><i>The police also found a pistol...</i></p>		<p>1 a: to come upon often accidentally : <u>encounter</u> b: to meet with (a particular reception) &lt;hoped to <i>find</i> favor&gt;</p>
2 [VERB] V n, V n n, V n for n, also V n for n to-inf	<p>If you <b>find</b> something that you need or want, you succeed in achieving or obtaining it.</p> <p><i>So far they have not found a way to fight the virus...</i></p>	<p>(=discover) to discover, especially where a thing or person is, either unexpectedly or by searching, or to discover where to obtain or how to achieve something: <i>I've just found a ten-pound note in my pocket.</i></p>	<p>to come upon by searching or effort: &lt;must <i>find</i> a suitable person for the job&gt;</p> <p>to discover by study or experiment: &lt;<i>find</i> an answer&gt;</p> <p>to discover by the intellect or the feelings : <u>experience</u> &lt;<i>find</i> much pleasure in your company&gt;</p>
3 [V-PASSIVE] be V-ed	<p>If something <b>is found</b> in a particular place or thing, it</p>	<p>3 <b>be found</b> to exist or be present somewhere: <i>Many plant and animal species are found only in the rainforests</i></p>	

No.	Collins	Cambridge	Merriam
	exists in that place.  <i>Fibre is found in cereal foods, beans, fruit and vegetables.</i>		
	<b>4</b> [VERB] V n -ing, V n -ed, V n prep/adv If you <b>find</b> someone or something in a particular situation, they are in that situation when you see them or come into contact with them.  <i>They found her walking alone and depressed on the beach...</i>	to become aware that something exists or has happened: <i>We came home to find <b>(that)</b> the cat had had kittens.</i>	
	<b>5</b> [VERB] V pron-refl prep/adv, V pron-refl -ing, V pron-refl adj If you <b>find yourself</b> doing something, you are doing it without deciding or intending to do it.  <i>It's not the first time that you've found yourself in this situation...</i>	to become aware that you are in a particular situation or place, or doing a particular thing, unintentionally: <i>He'll find himself with no friends at all if he carries on behaving this way.</i>	to perceive (oneself) to be in a certain place or condition  to bring (oneself) to a realization of one's powers or of one's proper sphere of activity:  <must help the student to <i>find</i> himself as an individual — N. M. Pusey>

No.	Collins	Cambridge	Merriam
6	[VERB] V that, V it adj to-inf, V n to-inf, V n		
	n		
	If you <b>find</b> that something is the case, you become aware of it or realize that it is the case.		
	<i>The two biologists found, to their surprise, that both groups of birds survived equally well...</i>		
	7 [VERB] be V-ed adj, V n adj	to make a judgment in a law court: <i>In a unanimous verdict, the jury found him <b>guilty/not guilty</b> of the murder.</i>	to determine and make a statement about: <find a verdict> <found her guilty>
	When a court or jury decides that a person on trial is guilty or innocent, you say that the person <b>has been found</b> guilty or not guilty.		to determine a case judicially by a verdict: <find for the defendant>
	<i>When they found us guilty, I just went blank.</i>		
	8 [VERB] V n adj, V it adj that, V n n	2 to think or feel a particular way about someone or something: <i>Do you find Clive difficult to talk to?</i>	
	You can use <b>find</b> to express your reaction to someone or something.		
	<i>We're sure you'll find it exciting!...</i>		

No.	Collins	Cambridge	Merriam
9	[VERB] V n in -ing, V n in n		
	<p>If you <b>find</b> a feeling such as pleasure or comfort <b>in</b> a particular thing or activity, you experience the feeling mentioned as a result of this thing or activity.</p> <p><i>How could anyone find pleasure in hunting and killing this beautiful creature?</i></p>		
	10 [VERB] V n, V n		to obtain by effort or management <find the time to study>
0	<p>If you <b>find</b> the time or money <b>to</b> do something, you succeed in making or obtaining enough time or money to do it.</p> <p><i>I was just finding more time to write music...</i></p>		
			= <u>provide</u> , <u>supply</u>
1			to furnish (room and board) especially as a condition of employment
			= <u>attain</u> , <u>reach</u> : <the bullet <i>found</i> its mark>

No.	Collins	Cambridge	Merriam
2			
3			to gain or regain the use or power of : <trying to <i>find</i> his tongue>

Table2: Senses of 'FIND OUT' in dictionaries

No.	Collins	Cambridge	Merriam
1	<p>[PHRASAL VERB] V P wh, V P that, V P n (not pron), V n P</p> <p>If you find something out, you learn something that you did not already know, especially by making a deliberate effort to do so.</p> <p>It makes you want to watch the next episode to find out what's going to happen...</p> <p>= discover</p>	<p>1 to get information about something because you want to know more about it, or to learn a fact or piece of information for the first time:</p> <p><i>How did you find out about the party?</i></p>	<p>1: to learn by study, observation, or search : <u>discover</u></p> <p><u>3</u> to discover, learn, or verify something &lt;I don't know, but I'll <i>find out</i> for you&gt;</p>
2	<p>[PHRASAL VERB] V n P</p>	<p>2 to discover that someone has done something wrong:</p>	<p>2 a: to catch in an offense (as a crime)</p>

No.	Collins	Cambridge	Merriam
	<p>If you <b>find</b> someone <b>out</b>, you discover that they have been doing something dishonest.</p> <p><i>Her face was so grave, I wondered for a moment if she'd found me out.</i></p>	<p><i>He lived in dread of being found out.</i></p>	<p>&lt;the culprits were soon <i>found out</i>&gt; 2b: to ascertain the true character or identity of &lt;the informer was <i>found out</i>&gt;</p>

## Appendix E

The senses of the five example PVs from Cambridge Phrasal Verbs Dictionary (McCarthy & Walter, 2006):

### 1. DRAW UP

to prepare something by writing it/ to move a piece of furniture near to something or someone/ to move your knees or legs closer to your body/ a vehicle arrives somewhere and stops/ to stand up very straight

### 2. BRING UP

look after a child/ to start to talk about a particular subject/ to vomit something

### 3. LOOK UP

to look at a book or computer to find a piece of information/ a situation is improving/ to visit someone

### 4. PICK UP

to lift something or someone by using hands/ to collect someone or something/ to get or buy something/ to buy something cheaply/ to learn a new skill or language by practising it/ speak or behave in a particular way/ to learn information from someone or something/ win a prize/ pay for something/ get infectious illness from someone/ a device receives signals/ to become aware of a smell/ to notice a mistake/ to earn money/ to make a place tidy/ to start something again/ to start talking to someone/ the police arrest someone/ to stand up again/ business, economy improves after a bad period/ the wind becomes stronger/ answer the phone/ a vehicle starts to go faster

### 5. GROW UP

gradually change to become an adult/ to begin to exist and then become bigger and more important



## Corpora, dictionaries and software

### Corpora

Bank of English. Available at: <http://www.titania.bham.ac.uk/>

Cambridge Learner Corpus (CLC). See:  
[http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/?site\\_locale=en\\_GB](http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/?site_locale=en_GB)

Chinese Learner English Corpus (CLEC). See:  
<http://lc.ust.hk/~center/conf2001/keynote/subsect4/link.html>

International Corpus of Learner English (ICLE). 2002. See:  
<http://www.uclouvain.be/en-cecl-icle.html>.

LOCNESS. See: <http://www.uclouvain.be/en-cecl-locness.html>. For the essay topics, see:  
<http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/LOCNESS1.htm>

Longman Learners' Corpus (LLC). Available at:  
<http://www.pearsonlongman.com/dictionaries/corpus/index.html>

Academia Sinica Balanced Corpus of Modern Chinese. See:  
<http://app.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh?ukey=-1824669696&qtype=0>

### Phrasal verbs dictionaries

Cambridge Phrasal Verbs Dictionary (CPVD), 2<sup>nd</sup> Ed. 2006. Cambridge: Cambridge University Press.

Collins CoBUILD Dictionary of Phrasal Verbs, 2nd Ed. 2002. London: Collins CoBUILD.

Macmillan Phrasal Verbs Plus (MPVP), 1<sup>st</sup> Ed. 2005. China: Macmillan.

Oxford Phrasal Verbs Dictionary, 2nd Ed. 2007. Oxford: Oxford University Press, USA.

### **Other dictionaries**

American Heritage Dictionary of Idioms, 2<sup>nd</sup> Ed. 2003. USA: Houghton Mifflin Harcourt.

Collins Cobuild Advanced Learner's English Dictionary (CALED). 5th Ed. 2006. London: Collins.

Cambridge Advanced Learner's Dictionary (CALD), 2nd Ed. 2005. Cambridge: Cambridge University Press. Available at:

<http://dictionary.cambridge.org/define.asp?key=101893&dict=CALD>

Merriam-Webster Collegiate Dictionary (MWCD), 11th Rev Ed., 2003. Springfield: HarperCollins. Available at: <http://www.merriam-webster.com/>

Phrasal Verb Demon, Available at:  
<http://www.phrasalverbdemon.com/dictionarya.htm>

Using English, Available at: <http://www.usingenglish.com/>

### **Software**

CLAWS. Available at: <http://ucrel.lancs.ac.uk/claws/>. For the tagging guide, see: [http://ucrel.lancs.ac.uk/bnc2sampler/guide\\_c7.htm](http://ucrel.lancs.ac.uk/bnc2sampler/guide_c7.htm). For the BNC2 POS-tagging Manual, see: <http://ucrel.lancs.ac.uk/bnc2/bnc2autotag.htm>

Log-likelihood Wizard. Available at: <http://ucrel.lancs.ac.uk/llwizard.html>

PowerGREP. 2008. Just Great Software Co. Ltd.

WordSmith4. 2004. Mike Scott. Lexical Analysis Software Ltd. Available at: <http://www.lexically.net/wordsmith/version4/index.htm>

## Bibliography

- Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on Computer* (pp. 132–142). London and New York: Longman.
- Aisenstadt, E. (1979). Collocability restrictions in dictionaries. In R. Hartmann (Ed.), *Dictionaries and Their Users* (pp. 71–74). Exeter: University of Exeter.
- Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL, Review of Applied Linguistics*, 53, 53–61.
- Altenberg, B. (2005). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 101–122). New York: Oxford University Press.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173–195.
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on Computer* (pp. 80–93). London and New York: Longman.
- Ammer, C. (2003). *The American Heritage Dictionary of Idioms* (2nd edition.). USA:

Houghton Mifflin Harcourt.

Armstrong, K. (2004). Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness, 13*(4), 213–224.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101–114.

Baldwin, T., & Villavicencio, A. (2002). Extracting the unextractable: A case study on verb-particles. In *The Sixth Conference on Computational Natural Language Learning (CoNLL 2002)* (pp. 98–104). Taipei, Taiwan.

Ball, C. (1994). Automated text analysis: Cautionary tales. *Literary and Linguistic Computing, 9*(4), 295–302.

Barkema, H. (1996). Idiomaticity and terminology: A multi-dimensional descriptive model. *Studia Linguistica, 50*(2), 125–160.

Barlow, M., & Kemmer, S. (1999). *Usage-based models of language*. Stanford, California: CSLI Publications, Center for the Study of Language and Information.

Benson, M., Benson, E., & Ilson, R. (1986). *The Bbi Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam: John Benjamins.

Bernardini, S. (2004). Corpora in the classroom. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15–36). Amsterdam: John Benjamins.

- Berry-Rogghe, G. (1974). Automatic identification of phrasal verbs. In J. L. Mitchell (Ed.), *Computers in the Humanities* (pp. 16–26). Edinburgh: Edinburgh University Press.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311.
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15(2), 169–189.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, England: Pearson ESL.
- Bolinger, D. (1971). *The Phrasal Verbs in English*. London: Harvard University Press.
- Burt, M. K., & Kiparsky, C. (1972). *The Gooficon: A Repair Manual for English*. Rowley, MA: Newbury House.
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216–236). New York: Routledge.
- Cadierno, T. (2004). Expressing motion events in a second language: A cognitive typological perspective. In M. Achard & S. Niemeier (Eds.), *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*

- (pp. 13–49). Berlin and New York: Walter de Gruyter.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar Book: An ESL/EFL Teacher's Course* (2nd ed.). Boston: Heinle & Heinle.
- Channell, J. (1994). *Vague Language*. Oxford: Open University Press.
- Charles, M. (2004). *The Construction of Stance: A Corpus-based Investigation of Two Contrasting Disciplines* (Unpublished PhD thesis). University of Birmingham, UK.
- Charles, M. (2011). Making concessions in academic writing: A corpus study of patterns and semantic sequences. In N. Groom & O. Mason (Eds.), *Proceedings of the Corpus Linguistics Conference 2011* (p. n.a.). UK: University of Birmingham.
- Chen, M. (2013). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, 18(3).
- Cheng, W., Greaves, C., Sinclair, J., & Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, 30(2), 236–252.
- Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, 11(4), 411–435.

- Claridge, C. (2000). *Multi-Word Verbs in Early Modern English: A Corpus-Based Study: A Corpus-based Study*. Amsterdam: Rodopi.
- Consigny, A. (2001). *The Semantics of Phrasal Verbs in English: A Data-Driven Study* (Unpublished PhD thesis). University of Liverpool, UK.
- Cook, G. (1998). The uses of reality: A reply to Ronald Cater. *ELT Journal*, 52(1), 57–64.
- Cornell, A. (1985). Realistic goals in teaching and learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching*, 23(4), 269–280.
- Courtney, R. (1983). *Longman Dictionary of Phrasal Verbs*. Harlow, England: Longman.
- Cowie, A. (1993). Getting to grips with phrasal verbs. *English Today*, 9(4), 38–41.
- Cowie, A. (1994). Phraseology. In R. E. Asher (Ed.), *The Encyclopedia of Language and Linguistics* (pp. 3168–71). Oxford: Oxford University Press.
- Cowie, A. (1998). Phraseological dictionaries: Some East-West Comparisons. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 209–228). New York: Oxford University Press.
- Cowie, A. (Ed.). (2005). *Phraseology: Theory, Analysis, and Applications*. New York: Oxford University Press.

Cowie, A., & Howarth, P. (1996). Phraseological competence and written proficiency.

*British Studies In Applied Linguistics*, 11, 80–93.

Cowie, A., & Mackin, R. (1993). *Oxford Dictionary of Phrasal Verbs*.

Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and

opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign*

*Language Learning and Teaching* (pp. 149–161). Amsterdam and

Philadelphia: John Benjamins.

Cross, J., & Papp, S. (2008). Creativity in the use of verb + noun combinations by

Chinese learners of English. In G. Gilquin, S. Papp, & M. B. Diez-Bedmar

(Eds.), *Linking up Contrastive and Learner Corpus Research* (pp. 57–81).

Amsterdam: Rodopi.

Cruse, D. (1986). *Lexical Semantics*. New York: Cambridge University Press.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A Case for contrastive

analysis. *Studies in Second Language Acquisition*, 7(1), 73–79.

Dalle, T. (1983). *Phrasal Verbs: The Effectiveness of Current Grammatical Theory in*

*Applied ESL and Some Pedagogical Implications* (PhD Thesis). The

University of Arizona, USA. Retrieved from ProQuest.

Danesi, M. (1995). Learning and teaching languages: The role of “conceptual fluency.”

*International Journal of Applied Linguistics*, 5(1), 3–20.

- Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly*, 33(1), 65–83.
- De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20)* (pp. 53–68). Freiburg im Breisgau: Rodopi, Amsterdam and Atlanta.
- De Cock, S. (2001). Recurrence and phraseology in learner and native corpora of speech and writing. *ICAME Journal*, 1(n.a.).
- De Cock, S. (2005). Phrasal verbs and learners: How well do they get on? In *Macmillan Phrasal Verbs Plus*. China: Macmillan.
- De Cock, S. (2011). Preferred patterns of use of positive and negative evaluative adjectives in native and learner speech: An ELT perspective. In A. Frankenberg-Garcia, L. Flowerdew, & G. Aston (Eds.), *New Trends in Corpora and Language Learning* (pp. 198–212). New York: Continuum International Publication Group.
- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automatic approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on Computer* (pp. 67–79). London and New York: Longman.

- De Rycker, T. (2005). Word power: Phrasal verbs and compounds. A cognitive approach. *ITL, Review of Applied Linguistics, 149-150*, 93–101.
- Dechert, H., & Lennon, P. (1989). Collocational blends of advanced second language learners: A preliminary analysis. In W. Oleksy (Ed.), *Contrastive Pragmatics* (pp. 131–168). Amsterdam and Philadelphia: John Benjamins.
- Dehé, N. (2002). *Particle Verbs in English: Syntax, information structure and intonation*. Amsterdam and Philadelphia: Benjamins.
- Dehé, N., Jackendoff, R. S., McIntyre, A., & Urban, S. (2002). *Verb-particle Explorations*. Berlin: Walter de Gruyter.
- Denes, P. B., & Pinson, E. N. (1963). *The Speech Chain*. Baltimore: Waverly Press.
- Diaz-Negrillo, A., & Fernandez-Dominguez, J. (2006). Error tagging systems for learner corpora. *RESLA, 19*, 83–102.
- Dirven, R. (2001). English phrasal verbs: Theory and didactic application. *Applied Cognitive Linguistics, 2*, 3–28.
- Dixon, R. (1982). The grammar of English phrasal verbs. *Australian Journal of Linguistics, 2*(1), 1–42.
- Durrant, P. (2008). *High-frequency collocations and second language learning* (Unpublished PhD thesis). University of Nottingham, UK.
- Ellis, N. (2008). Phraseology: The periphery and the heart of language. In F. Meunier

- & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 1–13). Amsterdam and Philadelphia: John Benjamins.
- Farrell, P. (2005). English verb-preposition construction: Contiguity and order. *Language*, 81(1), 96–137.
- Fillmore, C. (1978). On the organization of semantic information in the lexicon. *Parasession on the Lexicon*, 148–173.
- Firth, J. (1957a). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
- Firth, J. (1957b). *Studies in Linguistic Analysis*. Oxford: Blackwell Publishing.
- Fodor, J. A., & Katz, J. J. (1964). *The structure of language: Readings in the philosophy of language*. Englewood Cliffs: Prentice-Hall.
- Fontenelle, T. (2005). Discovering significant lexical functions in dictionary entries. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 189–207). New York: Oxford University Press.
- Frank, S., Bod, R., & Christiansen, M. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, 279(1747), 4522–4531. doi:10.1098/rspb.2012.1741
- Fraser, B. (1974). *The phrasal verb in English*. Cambridge: Harvard University Press.
- Fraser, B. (1977). *Verb-particle Combination in English* (Corr. ed.). USA: Academic Press Inc.

- Gallagher, A. (2006). *Explicit grammar instruction and the learning of phrasal verbs in English* (Unpublished PhD thesis). Northern Illinois University, USA.
- Gardner, D., & Davies, M. (2007). Pointing out Frequent Phrasal Verbs: A Corpus-Based Analysis. *TESOL Quarterly*, 41(2), 339–359.
- Gitsaki, C. (1996). *The development of ESL collocational knowledge* (Unpublished PhD thesis). University of Queensland, Center of Language Teaching and Research, Australia.
- Gleason, J. B., & Ratner, N. B. (1998). *Psycholinguistics*. Fort Worth: Harcourt Brace College Publishers.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Language in Contrast* (pp. 37–51). Lund: Lund University Press. Retrieved from <http://dial.academielouvain.be/handle/boreal:75847>
- Granger, S. (Ed.). (1998a). *Learner English on Computer*. London and New York: Longman.
- Granger, S. (1998b). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3–18). London and New York: Longman.
- Granger, S. (2004). Computer learner corpus research: Current status and future

- prospects. In C. U. & T. A. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123–145). Amsterdam and Atlanta: Rodopi.
- Granger, S. (2005a). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 145–160). New York: Oxford University Press.
- Granger, S. (2005b). Pushing back the limits of phraseology: How far can we go. In *Proceedings of the Phraseology 2005 Conference* (pp. 1–4). Louvain-la-Neuve: UCL.
- Granger, S. (2011). From phraseology to pedagogy: Challenges and prospects. In P. Uhrig & S. Schüller (Eds.), *Chunks in the Description of Language: A tribute to John Sinclair* (pp. 123–146). Berlin: Mouton de Gruyter.
- Granger, S., & Paquot, M. (2008). Distinguishing the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 27–50). Amsterdam: John Benjamins.
- Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on Computer*. London and New York: Longman.
- Granger, S., & Wynne, M. (2000). Optimising Measures of Lexical Variation in EFL Learner Corpora. In *Corpora Galore: Analyses and Techniques in Describing English: Papers from the Nineteenth International Conference on English*

- Language Research on Computerised Corpora* (Vol. 30, p. 249).  
Louvain-la-Neuve, Belgium: Rodopi.
- Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25(1), 38–61.
- Gries, S. (2003). *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum International Publishing Group.
- Gries, S. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 3–25). Amsterdam and Philadelphia: John Benjamins.
- Groom, N. (2007). *Phraseology and Epistemology in Humanities Writing* (Unpublished PhD thesis). University of Birmingham, UK.
- Gui, S., & Yang, H. (2002). *Chinese Learner English Corpus*. Shang-hai: Shang-hai Foreign Language Education Publisher.
- Guo, X. (2006). *Verbs in the Written English of Chinese Learners: A Corpus-based Comparison between Non-native Speakers and Native Speakers* (Unpublished PhD thesis). The University of Birmingham, UK.
- Häggglund, M. (2001). The use of English phrasal verbs in the (written and spoken) language of advanced Swedish learners. In *ICAME2001*. Louvain-la-Neuve, Belgium.

- Halliday, M. (1992). Language as system and language as instance: The corpus as a theoretical construct. In *Directions in Corpus Linguistics* (pp. 61–77). Berlin: Mouton.
- Hampe, B. (1997). Towards a solution of the phrasal verb puzzle: Considerations on some scattered pieces. *Lexicology-Berlin*, 3(2), 203–243.
- Handl, S. (2008). Essential collocations for learners of English: The role of collocational direction and weight. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 43–66). Amsterdam: John Benjamins.
- Hanl, M. (2012). Collocation Analysis. Retrieved from [http://linglit194.linglit.tu-darmstadt.de/linguisticsweb/bin/view/LinguisticsWeb/CollocationAnalysis#T\\_score](http://linglit194.linglit.tu-darmstadt.de/linguisticsweb/bin/view/LinguisticsWeb/CollocationAnalysis#T_score)
- Hannan, P. (1998). Particles and gravity: phrasal verbs with “up” and “down.” *Modern English Teacher*, 7(1), 21–27.
- Hausmann, F. (1989). Le dictionnaire de collocations. In F. Hausmann, H. Wiegand, & L. Zgusta (Eds.), *Worterbucher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie* (pp. 1010–1019). Berlin: de Gruyter.
- Herbst, T. (1996). What are collocations: Sandy beaches or false teeth? *English*

*Studies*, 77(4), 379–393. doi:10.1080/00138389608599038

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41(4), 667–683. doi:10.1016/j.pragma.2008.09.029

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some Implications for Language*. Tübingen: M. Niemeyer.

Howarth, P. A. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24–44. doi:10.1093/applin/19.1.24

Howarth, P. A. (2005). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 161–188). New York: Oxford University Press.

Hudson, J. (1998). *Perspectives on Fixedness*. Lund: Lund University Press.

Hudson, R. (2008). Word grammar, cognitive linguistics, and second language learning and teaching. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 89–113). New York: Routledge.

Hulstijn, J. H., & Marchena, E. (1989). *Avoidance: Grammatical or Semantic*

*Causes?*

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249–268.

Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13(3), 271–295.

Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners. In Aijmer (Ed.), *Corpora and language teaching* (Vol. 33, pp. 141–154). Amsterdam: John Benjamins.

Hunston, S., & Francis, G. (1999). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

James, C. (1998). *Errors in language learning and use*. London: Longman.

Johansson, S. (1975). *The Phrasal Verb in English*. Cambridge: Harvard University Press.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *Classroom Concordancing (English Language Research Journal 4)*, 1–16.

Johns, T. (1997). Kibbitzer. Retrieved March 7, 2009, from

<http://www.eisu2.bham.ac.uk/johnstf/revis024.htm>

Kaalep, H. J., & Muischnek, K. (2002). Using the text corpus to create a comprehensive list of phrasal verbs. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation* (pp. 101–105). Las Palmas: European Respiratory Soc Journal.

Kamakura, Y. (2007). Phraseology in a Learner Corpus Compared with the Phraseology of UK and US Students. In *Proceedings of the Corpus Linguistics Conference 2007* (p. n.a.). UK: University of Birmingham.

Kaszubski, P. (1998). Enhancing a writing textbook: A national perspective. In S. Granger (Ed.), *Learner English on Computer* (pp. 172–185). London and New York: Longman.

Kaszubski, P. (2000). *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach* (Unpublished PhD thesis). Adam Mickiewicz, Poznań.

Keck, C. (2004). Corpus linguistics and language teaching research: Bridging the gap. *Language Teaching Research*, 8(1), 83–109.

Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics Studies in honour of Jan Svartvik* (pp. 95–110). London: Longman.

- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 21–42). Amsterdam: John Benjamins.
- Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2), 135–155.
- Klein, E. (1995). *Living Up to the Challenge of “Phrasal Verbs”: A Didactic and Methodological Analysis of Verb + Particle Combinations in English*. Duisburg: L.A.U.D.
- König, E. (1973). *Englische Syntax*. Frankfurt: Athenäum Fischer.
- Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. *The Bilingual Lexicon*, 53481. Retrieved from [http://www.google.com/books?hl=en&lr=&id=Rpm9Eei4ZR8C&oi=fnd&pg=PA53&dq=kroll+accessing+conceptual+representations&ots=bzTWkU\\_5US&sig=3atExdi5Lr2IWacacJCCZWJummw](http://www.google.com/books?hl=en&lr=&id=Rpm9Eei4ZR8C&oi=fnd&pg=PA53&dq=kroll+accessing+conceptual+representations&ots=bzTWkU_5US&sig=3atExdi5Lr2IWacacJCCZWJummw)
- Kurtyka, A. (2001). Teaching English phrasal verbs: A cognitive approach. In M. Pütz, S. Niemeier, & R. Dirven (Eds.), *Applied Cognitive Linguistics* (pp. 29–54). Berlin: Walter de Gruyter.

- Lam, L. (2003). *Conceptual Fluency in Second Language Learning Chinese ESL Learners' Production of Out and Up Particles in English Verb Particle Constructions* (Unpublished PhD thesis). Chinese University of Hong Kong, Hong Kong.
- Laufer, B., & Eliasson, S. (1993). What causes avoidance of in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, 15, 35–48.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on Computer* (pp. xiv–xx). London and New York: Longman.
- Lennon, P. (1996). Getting “easy” verbs wrong at the advanced level. *IRAL. International Review of Applied Linguistics in Language Teaching*, 24(1), 23–36.
- Lennon, P. (2005). Lexical collocations and the teaching of lexical meaning. *Neusprachliche Mitteilungen Aus Wissenschaft Und Praxis*, 58(1-2), 18–24.
- Lesniewska, J. (2006). Collocations and second language use. *Studia Linguistica*, 95–105.
- Lewis, M. (1993). *The Lexical Approach*. London: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. Hove UK: Language Teaching Publications.

- Li, W., Zhang, X., Niu, C., Jiang, Y., & Srihari, R. (2003). An expert lexicon approach to identifying English phrasal verbs. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, 513–520.
- Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, 54(2), 193–226.
- Lindner, S. (1983). *A lexico-semantic analysis of English verb particle constructions with OUT and UP*. (Unpublished PhD thesis). University of Indiana, USA.
- Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, 45(4), 661–688. doi:10.5054/tq.2011.247707
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25–35. doi:10.1016/j.esp.2011.07.002
- Liu, T., & Shaw, P. M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL*, 39(3), 171–194.
- Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on Computer* (pp. 53–66). London and New York: Longman.

- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker (Ed.), *Text and Technology* (pp. 157–76). Amsterdam: John Benjamins.
- Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67(2), 184–198. doi:10.1093/elt/ccs100
- McArthur, T. (1979). The strange case of the English phrasal verb. *Zielsprache Englisch*, 9(3), 24–26.
- McArthur, T. (1989). The long-neglected phrasal verb. *English Today*, 5(02).
- McArthur, T. (Ed.). (1992). *The Oxford Companion to the English Language*. Oxford and New York: Oxford University Press, USA.
- McCarthy, M., & Walter, E. M. (Eds.). (2006). *Cambridge Phrasal Verbs Dictionary* (2nd edition.). Cambridge University Press.
- McEney, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- McGee, I. (2012). Should we teach semantic prosody awareness? *RELC Journal*, 43(2), 169–186.
- McIntosh, C. (2006). *Oxford Phrasal Verbs Dictionary for learners of English* (2nd edition.). Oxford: OUP Oxford.
- McKenny, J. A. (2006). *A Corpus-based Investigation of the Phraseology in Various*

- Genres of Written English with Applications to the Teaching of English for Academic Purposes* (Unpublished PhD thesis). University of Leeds (School of English), UK.
- Meunier, F., & Granger, S. (1998). Computer tools for the analysis of learner corpora. In S. Granger (Ed.), *Learner English on Computer* (pp. 19–37). London and New York: Longman.
- Meunier, F., & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam and Philadelphia: John Benjamins.
- Milton, J. (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In S. Granger (Ed.), *Learner English on Computer* (pp. 186–198). London and New York: Longman.
- Milton, J. (2000). *Elements of a written interlanguage: A computational and corpus-based study of institutional influences on the acquisition of English by Hong Kong Chinese students* (Unpublished PhD thesis). Lancaster University, UK.
- Mindt, D. (2000). *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
- Mish, F. (Ed.). (2003). *Merriam-Webster Collegiate Dictionary* (11th Revised

- edition.). Springfield: HarperCollins.
- Mollet, E., Wray, A., Fitzpatrick, T., Wray, N. R., & Wright, M. J. (2010). Choosing the best tools for comparative analyses of texts. *International Journal of Corpus Linguistics*, 15(4), 429–473. doi:10.1075/ijcl.15.4.01mol
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Moon, R. (2005). Metaphor and phrasal verbs. In *Macmillan Phrasal Verbs Plus*. China: Macmillan.
- Nation, I. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Neagu, M. (2007). English verb particles and their acquisition: A cognitive approach. *RESLA*, 20, 121–138.
- Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purpose*, 25, 217–234.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
- Nesselhauf, N. (2004). What are collocations? In D. Allerton, N. Nesselhauf, & P.

- Skandera (Eds.), *Phraseological Units: Basic Concepts and Their Application* (pp. 1–21). Basel: Schwabe.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Philadelphia: John Benjamins.
- Nesselhauf, N., & Römer, U. (2007). Lexical-grammatical patterns in spoken English: The case of the progressive with future time reference. *International Journal of Corpus Linguistics*, 12(3), 297–333.
- Oakes, M., P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Palmer, H. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101–119). Amsterdam: John Benjamins.
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149.
- Partington, A. (1998). *Patterns and Meanings*. Amsterdam: John Benjamins.
- Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131–

156.

- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–226). London: Longman.
- Potter, E. (2005). The syntactic behaviour of phrasal verbs. In *Macmillan Phrasal Verbs Plus*. China: Macmillan.
- Poulsen, S. (2005). *Collocations as a language resource. A functional and cognitive study in English phraseology* (Unpublished PhD thesis). University of Southern Denmark, Denmark.
- Pozdnyakova, O. V., & Gunina, N. A. (2011). Teaching phrasal verbs. *Problems of Contemporary Science and Practice, n.a.*, 357–360.
- Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26(81), 114.
- Prodromou, L. (2003). Idiomaticity and the non-native speaker. *English Today*, 19(2), 42–48.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Rayson, P. (2010). Log-likelihood Wizard. Retrieved from <http://ucrel.lancs.ac.uk/llwizard.html>
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In

- Proceedings of the workshop on comparing corpora* (pp. 1–6). Hong Kong: Hong Kong University of Science and Technology.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 209–227). Amsterdam and Philadelphia: John Benjamins.
- Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on Computer* (pp. 41–52). London and New York: Longman.
- Rundell, M., Shovel, M., & Fox, G. (2005). *Macmillan Phrasal Verbs Plus* (1st ed.). China: Macmillan.
- Römer, U. (2005). *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts And Didactics*. Amsterdam: John Benjamins.
- Saeed, J. I. (2000). *Semantics: Introducing Linguistics*. Oxford: Blackwell.
- Sawyer, J. (1999). *Verb adverb and verb particle constructions: Their syntax and acquisition*. (PhD Thesis). Boston University, USA. Retrieved from UMI Dissertation Services.
- Sawyer, J. (2000). Comments on Clayton M. Darwin and Loretta S. Gray's "Going

- after the Phrasal Verb: An Alternative Approach to Classification”. A Reader Reacts. *TESOL Quarterly*, 34(1), 151–159. doi:10.2307/3588100
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequence: acquisition, processing and use*. Amsterdam: John Benjamins.
- Schneider, E. (2004). How to trace structural nativization: Particle verbs in world English. *World Englishes*, 23(2), 227–249.
- Scott, M. (2004). WordSmith (Version 4.0). Retrieved from [www.lexically.net](http://www.lexically.net)
- Shei, C. (2005). Fixedness in genre-specific language and intercultural differences. *International Journal of Corpus Linguistics*, 10(2), 199–225.
- Shei, C. (2008). Discovering the hidden treasure on the Internet: Using Google to uncover the veil of phraseology. *Computer Assisted Language Learning*, 21(1), 67–85.
- Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal*, 44(2), 144–152.
- Siepmann, D. (2006). Collocation, Colligation and Encoding Dictionaries. Part II: Lexicographical Aspects. *International Journal of Lexicography*, 19(1).
- Sinclair, J. (1966). Beginning the study of lexis. In C. Bazell, J. Catford, M. Halliday, & R. Robins (Eds.), *In Memory of J.R.Firth* (pp. 410–430). London: Longman.

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

Sinclair, J. (1996). The search for units of meaning. *TEXTUS*, IX(1), 75–106.

Sinclair, J. (1999). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics*. Amsterdam: John Benjamins.

Sinclair, J. (2004a). *English Grammar*. London: HarperCollins.

Sinclair, J. (2004b). *Trust the text : language, corpus and discourse*. London and New York: Routledge.

Sinclair, J. (2006). *Collins COBUILD Advanced Learner's English Dictionary* (5th ed.). London: Collins.

Sinclair, J. (2007a). *Collocation reviewed*. Tuscan Word Centre, Italy.

Sinclair, J. (2007b). *Defining the definiendum-new*. Tuscan Word Centre, Italy.

Sinclair, J., Hanks, P., & Moon, R. (Eds.). (2002). *Collins Cobuild Dictionary of Phrasal Verbs* (2nd edition.). London: Collins.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140–158). London: Longman. Retrieved from [http://rdues.bcu.ac.uk/publ/Lexical\\_Syllabus.pdf](http://rdues.bcu.ac.uk/publ/Lexical_Syllabus.pdf)

Singleton, D. (1999). *Exploring the Second Language Mental Lexicon*. Cambridge:

Cambridge University Press.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching*, 45(2), 119–139.

Sjöholm, K. (1995). *The Influence of Crosslinguistic, Semantic, and Input Factors on the Acquisition of English Phrasal Verbs: A Comparison between Finnish and Swedish Learners at an Intermediate and Advanced Level*. Åbo: Åbo Akademi University Press.

Smidowicz, V. (1997). Phrasal Verbs...and How to Put Up with Them. *Zielsprache Englisch*, 27(4), 13–14.

Sonaiya, R. (1991). Vocabulary Acquisition as a Process of Continuous Lexical Disambiguation. *International Review of Applied Linguistics in Language Teaching*, 29(4), 273–284.

Sroka, K. (1972). *The Syntax of English Phrasal Verbs*. Paris:Mouton: The Hague.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23–55.

Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell Publishers.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.

- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7, 215–244.  
doi:doi:10.1075/ijcl.7.2.04stu
- Stubbs, M. (2009). John Sinclair (1933–2007): The Search for Units of Meaning-Sinclair on Empirical Semantics. *Applied Linguistics*, 30(1), 115–137.
- Sweetser, E. (1990). *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.
- Televnaja, J. (2004). *Lexical acquisition and semantic representation of English phrasal verbs in ontological semantics* (Unpublished PhD thesis). Purdue University, USA.
- Teliya, V., Bragina, N., Oparina, N., & Sandomirskaya, I. (2005). Phraseology as a Language of Culture: Its Role in the Representation of a Collective Mentality. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Application* (pp. 161–186). New York: Oxford University Press.
- Thibeau, T. J. (1999). *English prepositions in phrasal verbs: A study in second language acquisition* (Unpublished PhD thesis). The University of Arizona, USA.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work* (p. 223). John Benjamins

Publishing.

Villavicencio, A. (2003). Verb-particle constructions and lexical resources. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment* (Vol. 18, pp. 57–64). n.a.: Association for Computational Linguistics.

Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students* (Unpublished PhD thesis). Albert-Ludwigs-Universität Freiburg., Freiburg.

Waller, T. (1993). Characteristics of near-native proficiency in writing. *Near-Native Proficiency in English*, 2, 183–293.

Walter, E. (2005). *Cambridge Advanced Learner's Dictionary* (2nd ed.). Cambridge: Cambridge University Press.

Warren, B. (2005). A model of idiomaticity. *Nordic Journal of English Studies*, 4(1), 35–54.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2), 180–205.

White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, 96(3), 419–438.

Wible, D. (2008). Multiword expressions and the digital turn. In F. Meunier & S.

- Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 163–184). Amsterdam: John Benjamins.
- Wiktorsson, M. (2002, March 5). Register differences between prefabs in native and EFL English. Retrieved from <http://www.englund.lu.se/research/workingpapers/Volume-1-2001.phtml>
- Willis, D. (1990). *The Lexical Syllabus: An New Approach to Language Teaching*. London: Collins ELT.
- Wolter, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, 27(4), 741 – 747. doi:10.1093/applin/aml036
- Wong-Fillmore, L. (1976). *The second time around: Cognition and social strategies in second language acquisition* (Unpublished PhD thesis). Stanford University, USA.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. London: Continuum.
- Wyss, R. (2003). Putting Phrasal Verbs into Perspective. *TESOL Journal*, 12(1), 37–38.

- Xiao, Z., & McEnery, T. (2006, July). Corpora and language education. Retrieved from [http://www.lancs.ac.uk/postgrad/xiaoz/papers/corpora\\_teaching.doc](http://www.lancs.ac.uk/postgrad/xiaoz/papers/corpora_teaching.doc)
- Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *Tesol Quarterly*, 44(2), 250–273.
- Yorio, C. A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly*, XIV, 433–442.
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the Lifespan: Aspects of Acquisition, Maturity, and Loss* (pp. 55–72). Cambridge: Cambridge University Press.
- Yoshitomi, A. (2006). The Use of Phrasal Verbs by Japanese Learners of English--Implications from Story Telling Data. In *Readings in Second Language Pedagogy and Second Language Acquisition: In Japanese Context* (pp. 201–225). Amsterdam: John Benjamins.

---

<sup>i</sup> “AGENT is the initiator of some action...; PATIENT is the entity undergoing the effect of some action...; THEME is the entity which is moved by an action...” (Saeed, 2000:140).

<sup>ii</sup> Following Granger and Rayson (1998:121), log-likelihood values are used in this study instead of the chi-square, because the chi-square has problems in comparing two corpora of substantially different sizes.

<sup>iii</sup> The very frequent PVs which have frequencies over 100 are not considered. I then selected these five example PVs starting from the candidates which occur less than 100 in descending order (the minimum frequency is set to at least 10 times).

<sup>iv</sup> American slang used in the mid-1900s to mean ‘appear, become prominent’ (from The American Heritage Dictionary of Idioms (2003:83)).

<sup>v</sup> The collocates were retrieved in a 4:4 span (see: [http://www.titania.bham.ac.uk/docs/svenguide.html#The Collocations Option](http://www.titania.bham.ac.uk/docs/svenguide.html#The_Collocations_Option) ).

<sup>vi</sup> This was done by taking out the lines of *trying to find out* by using regular expressions.

<sup>vii</sup> The 606 instances comprise 280 times of ‘*time goes on*’, 283 times of ‘*time went on*’ and 43 times of ‘*time has/had gone on*’.

<sup>viii</sup> I consulted the ‘Academia Sinica Balanced Corpus of Modern Chinese’ with the Chinese synonymous gloss of ‘to go on’ and found 1144 instances. 100 random cases were then extracted and examined manually. Nearly one third of the 100 examples contain a ‘model-like’ verb.

<sup>ix</sup> This category contains two examples of ‘*a lot of*’, which was excluded from the figure.

<sup>x</sup> Although *downfall*, *collapse* and *death* seem to suggest negative consequences, looking into the complete concordance lines shows that this assumption is not always true. Some instances are actually referring to the downfall, collapse or death of something undesired, as in: *What about Wojtyla's individual contribution as the sole Polish Pope in history, invariably credited with helping to bring about the collapse of the Evil Empire itself, and...[BoE].*